California's Offical Redistricting Database–2021

Kenneth F. McCue, Ph.D.

Contents

Summary

Redistricting in California requires the availability of a database that includes counts of the total and voting age population, race and ethnicity, voter registration, and electoral results to ensure compliance with federal and state law. The Federal Voting Rights Act (FVRA) is the most important of these, and caselaw interpreting that Act lays out criteria for determining whether a districting plan complies with the FVRA. To determine whether a plan will comply with the FVRA, it is necessary to look at voter registration and historic electoral returns.

Other legal criteria that state plans must satisfy in California were laid out in Propositions 11 (2008) and 20 (2010). In addition, the California Citizens Redistricting Commission (CRC) established under the procedures in Propositions 11 and 20 has wide latitude in making decisions on redistricting plans and requires data about communities of interest, defined by economic and social characteristics.

California law (Elections Code, section 21003) also requires the adjustment of the decennial census data which forms the baseline of the redistricting dataset, such that data about persons incarcerated and enumerated in a state correctional facility must be reassigned to their last known residential address. Additionally, data about individuals incarcerated in a federal correctional facility must be excluded. The CRC must determine whether to use the adjusted dataset.

Local jurisdictions including cities and counties are mandated to use the official redistricting database for their redistrictings. (Assem. Bill 849, 2019-2020, ch. 557, 2019 Cal. Stat)

The primary conceptual difficulties in creating this database are caused by the fact that the component datasets are provided in aggregate form and are from different and uncoordinated sources. Aggregate means that the data are not reported at the level of the individuals, but rather as an agglomeration over a (usually) contiguous geographic area. The different sources are California's 58 county registrars of voters offices and the United States Census Bureau. If all the data from these sources were available at the level of the individual, the creation of the database would be a relatively simple matter, but most data are not available at the individual level.

Breaking down the data into component sets, the first type is census data. Census data are collected at the individual level, but are only reported at certain levels of aggregation, the smallest level being the census block (census tracts are composed of census blocks, and counties are made up of census tracts). The census block is typically (in urban areas) a city block, though there are exceptions to this. The census dataset for redistricting (P.L. 94-171) is based at the level of the census block and has data on population, ethnicity, age, and housing.

The second type of data is registered voter data. Just as with the census data, although these data are collected at the individual level they are not reported

that way, and just as the Census Bureau does not report individual data, there are no data reported at the level of the individual in the statewide redistricting database. The registration data collected on individuals is maintained by the California county registrars of voters and county clerk offices in the 58 counties and is reported to the California Secretary of State's office (and is referred to as the statewide registered voter file). Data from this file that are used for the statewide database include the voter party preference (as declared when registering), date of birth (used to derive age), surname (used in ethnic surname matching), address (used for placing the voter into that voter's census block) and voter history. These collected statistics are reported at the census block level. In particular, the names of individual voters are not included in the statewide database.

The third type of data is election results, which is organized by election result precincts, called SOV (Statement of Vote) precincts. These data are collected from each of the 58 county elections offices and are used to analyze racially polarized voting under the VRA. SOV precincts are made up of one or more registered voter precincts, and so these precincts are sometimes called consolidated precincts by the California county registrars of voters and county clerk offices. These data are also collected individually (as are the census and registered voter information), but are only reported at the SOV precinct level. The data available are the votes on statewide constitutional offices, state legislative races, federal races, and statewide propositions.

The fourth type of data is on individuals incarcerated in state correctional facilities. The California Department of Corrections and Rehabilitation (CDCR) provided a list of persons incarcerated at state correctional facilities in the state of California and the last address of those persons prior to incarceration, if this address was available. The 2020 census enumerated incarcerated persons at the correctional facility in which they were incarcerated at the time of the scheduled census enumeration, so the CDCR data are also correlated with the 2020 census data from the census blocks of the correctional facilities.

These four types of data are thus reported by different geographic units. For redistricting purposes, the registration and electoral data need to be placed into census blocks. Dealing with the registration data first, these data are placed into the census blocks by several methods. The first method is geocoding, that is, determining the census block of an address by using an equivalency table between an address range and a census block. For example, in a particular city, 100 E Elm Street might be in census block 1003, which is itself within tract 203.01. The name for the database where these equivalencies can be found is called the TIGER system, where TIGER stands for Topologically Integrated Geographic Encoding and Referencing system, which contains not only these address- to-census block equivalencies, but also a complete topographic representation of all census geography.

Much as the Census Bureau can then take information collected from individuals and report it at the census block level, the data which are geocoded from the

registered voter file can then be aggregated to the census block level and reported at that level (say, the number of voters with Spanish surnames). Not all addresses in the registered voter file can be geocoded, however. Errors can occur due to the incompleteness of the Census Bureau equivalency tables, differences in spellings of street names between the two systems, differences in representations of street types or directions, or differences for a particular address range between the block given in the TIGER file for that street range and the block at which the Census Bureau actually reported the information.

These addresses unassigned through geocoding can be handled by a variety of methods. Differences in street names between the registered voter file and TIGER files can be adjusted with synonym tables; differences in street types can be handled by examining whether there are unique street types within both the TIGER and registered voter file; and so forth. When these methods do not result in an assignment, a geographic conversion is used to allow the conversion of registration precincts to census geography. This is done by creating map overlays between the geographies of the registered voter precincts and those of the census geography. Since the geographic representation of the precincts and Census Bureau come from different sources, the reconciliation of the two types of geography is a time-consuming process, and at times registered voter precincts must be combined to allow accurate mapping. These precinct overlays are then used to assign registered voters' addresses when geocoding does not work.

The electoral results can now be allocated to the census blocks with the following method. Every SOV precinct can be associated with one or more registration precincts. Since every registered voter's address in a precinct has now been assigned to a census block, it is now possible to assign, for every voter, a portion of the vote for any particular candidate or proposition. The assignment of this portion is done through the use of a common statistical practice known as ecological inference, which has been accepted widely in both litigation and academic research. This statistical procedure gives an overall rate of support for any particular candidate or proposition, and this overall rate is then adjusted for the actual results in a particular precinct. These adjusted levels of support are then used to allocate election results to the individual voters, and these individual vote propensities are then aggregated up to the census block level to produce electoral results at that level.

While this is the most complete and accurate methodology for merging data to the census geography, there are times when it is desirable to add other data to the database, such as the citizen voting age population (CVAP) data. In this case, the data are in the 2010 census geography and must be converted into the 2020 census geography. The CVAP data are reported at the 2010 census block group level (there are typically 10 to 20 blocks in a block group), and furthermore, they are estimated from the ACS surveys, which are samples rather than complete enumeration.

For redistricting purposes, census and CVAP data must also be adjusted to reallocate data from incarcerated persons back to their last known residential

addresses. A similar geocoding process to the one described above is used to remove data from census blocks containing state correctional facilities, and to reallocate that data to census blocks containing the last known addresses of incarcerated persons, or to the most restrictive geography available if full geocoding is not possible.

Data Sources

Census Data

Census data are data collected every ten years by the Census Bureau (this collection is mandated by the US Constitution). Both state and federal law requires that districts be of equal population. The Voting Rights Act, an act of legislation by Congress, contains additional requirements dealing with how new district lines affect minority populations which have been historically discriminated against.

P.L. 94-171 is the law under which population data are provided to the states for redistricting. The data in this law contain counts on population, both by age and by ethnicity and race. These data are reported by census block. A census block is typically a city block in urban areas (though not always--this is discussed in the section on geocoding) and in rural areas it can comprise a much larger geography. This is illustrated in , which is taken from the P.L. 94-171 documentation (it is referred to as Figure 3 in that documentation). The Census Bureau divides counties into census tracts which are made up of block groups which contain blocks (note that block groups are used in the CVAP [Citizen Voting-Age Population], which is discussed in the section on CVAP breakdown).

It should be noted that this is an idealized figure, as sometimes census blocks are not equal to city blocks. For example, in , the dark orange shaded part of the figure is a 2010 census block. It is also a city street rather than a city block (a full description of this particular city block and the type of problems it presents is in Appendix "Potential Geocoding Problems"). The same remained true for 2020 census blocks.

The P.L.94-171 data are reported in six tables and are documented in the 2020 Census Redistricting Data (Public Law 94-171) Summary File Technical Documentation prepared by the U.S. Census Bureau, 2021. The P.L. 94-171 file's six tables (one of which relates to housing) are displayed in table 1. The population tables are P1 through P5. In P1, individuals are broken down by race; in P2, individuals are broken down by Latino versus non-Latino by race; in P3, 18 and older individuals are broken down by race; in P4, 18 and older individuals are broken down by Latino versus non-Latino by race; and P5 contains counts for individuals in group quarters. This last table is used for adjustments to the P.L. 94-171 data (see section "Data from Incarcerated Persons and P.L. 94-171 Adjustments"). Only table P2 and P4 are adjusted, for reasons explained in that section.

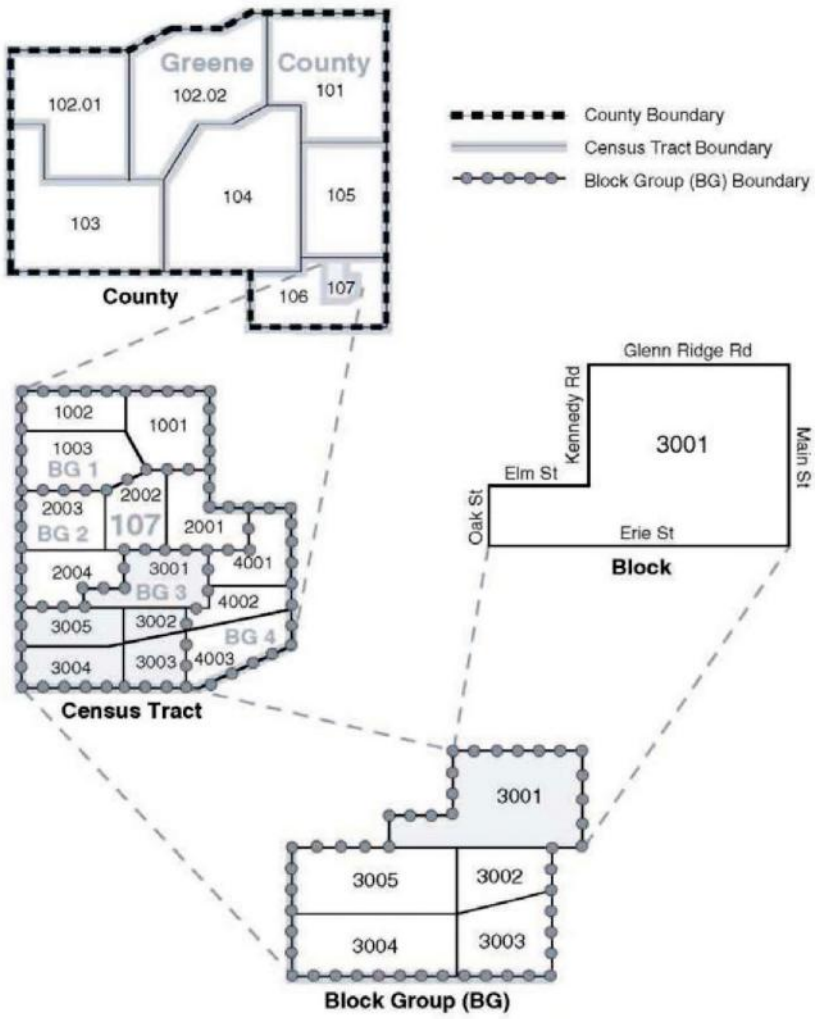


Figure 1: Geographic Relationships--Small Area Statistical Entities, County-Census Tract-Block Group-Block

There are various requirements for drawing districts. One is the requirement of nearly equal populations in each district. Thus each district should be near an "ideal" number, that is the total number of people in a political entity divided by the number of districts. In California, for example, with a California adjusted PL94-171 2020 total population of 39,523,437, the ideal population for each Assembly district is 494,041.5875 people. The race data are also used, in conjunction with political and registration data, to determine compliance with the Federal Voting Rights Act. In addition, a special tabulation of Citizen Voting Age Population (CVAP) is merged into the dataset. This dataset has unique

Figure 2: A 2010 Census Block Which is Not a City Block

Table 1: P.L. 94-171 Tables

P1	Race
P2	Hispanic or Latino, and not Hispanic or Latino by Race
P3	Race for the Population 18 Years and Over
P4	Hispanic or Latino, and not Hispanic or Latino by Race for the Population 18 Years and Over
P5	Group Quarters Population by Major Group Quarters Type
H1	Occupancy Status (Housing)

characteristics and is described under the section "CVAP Data". Additional technical documentation on the Public Law 94-171 Summary File can be found on the Census Redistricting Data page.

Registered Voter Data

Registered voter data are maintained by California county registrars of voters and county clerk offices on individuals who are registered to vote. All data reported under registered voter statistics are derive from the registered voter data, either directly or with additional procedures applied (described below). There are various legal requirements for voting and these legal requirements determine the information the registrar collects. Individuals are legally required to live in the districts they are voting in, so the registrar maintains the home address of every voter. Voters must also be 18 years old before they are eligible

to vote in an election, so the registrar maintains the date of birth. Information on the voter's declared party preference is kept so that the registrar may ensure that only voters who have declared a preference for a party may vote on party primary contests, as determined by political party rules. The voters' date of registration is also kept, as voters may reregister at any time (to change party preference, for example). Additionally, the voter's name is included on this record, which will be used for ethnic coding.

Statistics derived from the registered voter file consist of the following categorizations, either singly or combined: party preference, age, ethnicity, gender, and time since registration. These derived statistics are calculated at the level of the individual, and then reported at different levels of geographic aggregation: the census block, the registration precinct, and the aggregated voting precinct. Each categorization consists of two or more sub-categories, and for each categorization an individual is assigned to one and only one sub-category, with the value one for the assigned sub-categorization and zero for all others. This restriction ensures that aggregation from any geographic unit of the statistics to any other level (for example, from registration precinct to county) returns the correct values. In addition to these single categorizations, some combined categorizations (such as ethnicity by party preference) are reported, as described below.

Transformation of the registered voter file information into categorizations varies depending upon the information in the registered voter files. For the party preference categorization, subcategorizations are defined by the Republican, Democratic, Declined to State (also referred to as No Party Preference), Miscellaneous, and the "minor" parties with official recognition (California has a minimum registration number requirement for listing as an officially recognized party). The conversion is made by giving a voter with a Democratic preference an assignment to the Democratic sub-categorization, a voter with a Republican preference an assignment to the Republican sub-categorization, and so forth. Non-recognized parties are put into a miscellaneous sub-category and Declined to State voters are put into the Declined to State sub-categorization.

Gender is often not specified and so voter records which do not have a specified gender are assigned one through name matching. Because the concept of matching names is used also to provide an indication of ethnic or racial membership, gender is a good example of how the process of name matching occurs. Name matching requires a reference list to which the name can be matched and a characterization for the name returned. For gender, a "dictionary" (as lists of names with characterizations are typically called) was created by taking all names from the various counties' registered voter files that had a gender specified, and, if a name appeared more often as a female than a male, assigning that name as a female, and if it appeared less often assigning that name as a male.

Surname dictionaries for the creation of ethnic statistics work similarly. While the creation of a gender dictionary is relatively simple, ethnic surname dictionaries are much more complicated and require considerable effort to develop. The statewide database uses two such surname dictionaries, an Asian and a

Latino, which are published by the Census Bureau (a third dictionary, that of "Jewish" surnames, was developed by a political consulting firm in the late 1970s). Documentation on the Asian surname dictionary can be found in "Asian American ethnic identification by surname," Diane S. Lauderdale & Bert Kestenbaum, Population Research and Policy Review 19: 283-300, 2000. The Spanish-surnamed dictionary is referred to as the Passel-Word (PW) Spanish surname list, and is documented at: https://www.census.gov/library/working-papers/1993/demo/POP-twps0004.html

Matching with these surname dictionaries is similar in concept for that of a gender dictionary, but somewhat more complicated. For the Asian surname dictionary, for example, there are six different ethnicities a name can be assigned to (Korean, Japanese, Chinese, Indian, Vietnamese, and Filipino). For the Spanish surname dictionary, aside from direct last name to last name matching, various manipulations are made to surnames in order to fit them into the list.

The five basic categorizations of partisan affiliation, age, ethnicity, gender, and time since registration can be combined into multiple categorizations. For example, the various ethnic/ racial classifications are combined with partisan affiliation to create ethnicity by party. Age, gender, and partisan affiliation are combined to create counts by, say, Democratic males between 45 and 54 (inclusive). Partisan affiliations at the time of registration are used to create variables such as the number of Declined to State voters who registered after the second to last general election. Once again, when these multiple categorizations are created, an assignment of a voter is made to one and only one of the sub-categorizations of the multiple categorization set. For example, if a voter is a Democratic Chinese male, then only the sub-categorization Democratic Chinese would be assigned a value and all other sub-categorizations would be zero.

The complete list of variables derived from the registered voter lists is given in appendix "Registration data".

Election Data

Election data in the statewide database include results from 22 statewide elections held in California over the last decade. These elections are the general elections in 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016, 2018, and 2020; the primary elections in 2004, 2006, 2008, 2010, 2012, 2014, 2016, 2018, and 2020; the presidential primary in 2008; the gubernatorial recall election in 2003; and the special election in 2005. General elections are referred to by the letter g and the last two digits of the year it occurred (so g02, g04, g06, g08, g10, g12, g14, g16, g18, g20), primary elections similarly (so p04, p06, p08, p10, p12, p14, p16, p18, p20), and the other elections have an s and the year (so s03 for the gubernatorial recall, s05 for the special statewide election called by Governor Schwarzenegger, and s08 for the presidential-only primary in February of 2008).

The contests contained in the database consist of those pertaining to statewide constitutional offices, federal offices, state and federal legislative districts, and

statewide propositions, initiatives, and referenda. Merging election results to census geography is complicated and is described in its own section ("Disaggregation of Precinct Voting Results into Census Geography").

Election results are reported by statement of vote (SOV) precinct. The nature of SOV precincts has changed over the last decade with respect to absentee voting (also known as mail voting), and an understanding of these changes is important for knowledge of how to use the database. One important trend in California has been the rise of absentee voting, so that in current elections, an absentee participation rate of fifty percent or above is not atypical. In recognition of this fact, the Legislature passed a law before the 2008 elections requiring that absentee voters in a particular physical precinct must also have their electoral results reported in a separate precinct associated with that physical precinct. Thus for every voting precinct with polling place results, there is an additional precinct with voting results for those voters in that voting precinct who voted absentee. Before 2008, these absentee voters were often lumped together into common "ballot group" precincts, that is, a precinct reporting the election results of a group of voters who all share the same type of ballot and consisting of absentee votes for multiple precincts. Thus for counties which did not have an associated absentee precincts for each polling place precinct, SOV data for absentees is not allocated to geography.

Although, some counties did report absentee results by precincts which corresponded with polling place precincts before 2008, the largest counties in the state (Los Angeles, Orange, San Diego) did not. This has an effect on the allocation of voting results to census geography, which will be discussed in the section on "Disaggregation of Precinct Voting Results to Census Geography". But as a consequence, the vote totals for state elections will be a poor match for these elections previous to 2008, as these ballot group absentees are not used in the disaggregation. Even the vote totals past 2008 will not be an exact match since absentee precincts still exist. A case would be federal voters, since federal law allows one to register to vote for federal elections seven days before the election, rather than the 15 days required in California. But these types of absentee precincts have many fewer voters than in the ballot group precincts and the totals will be much closer for post-2006 elections than pre-2008 elections.

Creating a Common Geography

It is necessary to put the census, registration, and electoral data into a common unit of geography in order to use the database. These uses are varied: for example, census data at the census block level are used for redistricting, and election data at the voting precinct level are used for Voting Rights Act analysis along with ethnicity at the registration precinct level. And of course the Voting Rights Act analyses are also used in redistricting, so there must be a way of associating the results of the Voting Rights Act analysis to the census block.

This association of census, registration, and election results into units which can be used for redistricting is a central function of the statewide database. For redistricting purposes, this association is best done at the census block level, as various court cases have held that districts must meet certain population deviation requirements. These requirements are most easily met by performing redistricting by census blocks (and their aggregates, block groups and tracts), and associating other data, such as those used for Voting Rights Act compliance, to the census blocks.

The simplest manner of creating a dataset at the census block level would be to overlay the registration and voting precinct geographies on the census data. As the following figure makes clear, this will not produce a very accurate dataset. The figure shows a census block illustrated in dark orange and its intersection with various registration precincts (demarcated by red lines). As can be seen, any overlay of registration geography to census geography would result in large inaccuracies.

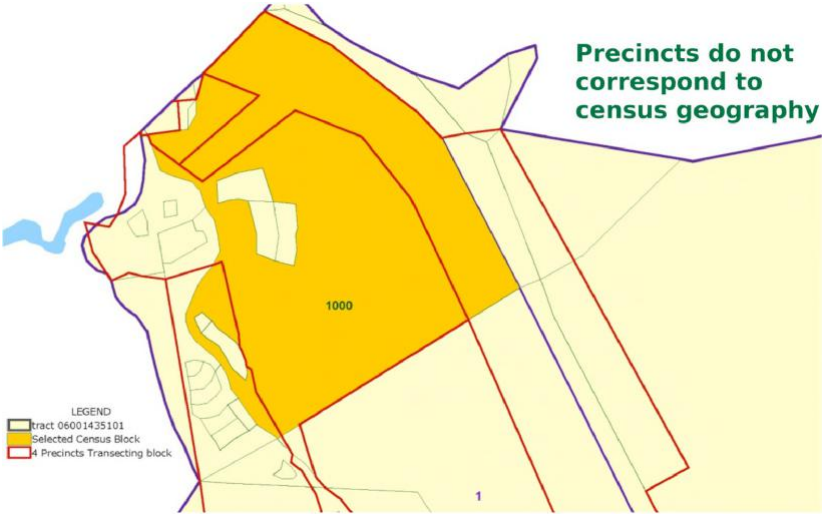


Figure 3: Precincts and Geography

To overcome the problem of using geographic overlays to create a merged dataset of census, registration and election data, then, the individual is treated as the unit of association. What all three types of geography have in common is that they describe individuals. For the census data, individuals are described only at the block level, and for electoral data, they are described at the voting precinct level. Only for the registration data are data on the individual available. The following sections explain, given these restrictions, how the merged dataset of census, registration and election data are constructed. The primary methods used are geocoding for registration data and ecological regression for election data, which are described in detail below.

Placing of Precinct Registered Voters into Census Geography

Registered voters are placed in census blocks by the process known as geocoding. Geocoding is the process of using address files which cross-reference addresses to census geography to assign the address of a registered voter to a particular census block. This assignment is never perfect, and so there are registered voters who are unassigned through geocoding who must still be assigned to a census block.

This is done in the following manner. First, equivalencies of precincts to census geography are created. This process involves taking precinct geography and overlaying it on the census geography. This overlaying is done on Geographic Information Systems (GIS) and adjustments to the two mapping systems due to different coordinate systems, different base maps, and various other factors are made by hand.

This overlay between precinct geography and census geography provides a list, for every precinct, of the census blocks which totally or partially fall into the precinct. These lists can then be used to assign registered voters to a census block, for registered voters who were not assigned through geocoding. This assignment is done through the criteria of attempting to equate for all blocks, as closely as possible, the ratio of registration to population.

This can be reformulated as a constrained optimization problem as follows. Let b be a census block in a unit of geography being balanced (one way to think of this problem is as balancing the ratios of registration to population among the blocks), and let there be B blocks. Let s be a precinct, and let a_{sb} be the assigned registration through geocoding of precinct s in block b, and let u_{sb} be the unassigned registration in precinct s and block b. The u_{sb} are the unknowns to be solved for, and they satisfy

$$r_s = \sum_{b \in r} (a_{sb} + u_{sb}),\ u_{sb} > 0,$$

where r_s is the registration in precinct r. This says simply that all of the registration, either assigned through geocoding or through equating ratios, must sum to the total registration in the precinct. Given this constraint, then, the optimization problem becomes to find a set of u_{rb} such that

$$\frac{\sum_{r \in b}(a_{rb} + u_{rb})}{p_b} = \frac{\sum_{r \in c}(a_{rc} + u_{rc})}{p_c},\ \text{for all } b, c \in B$$

Such a set may not exist but then a formulation such as

$$\min \sum_{b,c} \left[\frac{\sum_{r \in b}(a_{rb} + u_{rb})}{p_b} - \frac{\sum_{r \in c}(a_{rc} + u_{rc})}{p_c} \right]^2,\ \text{for all } b, c \in B$$

can be adopted.

Table 2: Assignment of Unassigned Registration to Equate Registration/Population Ratios

	Assigned by Geocoding			Unassigned Allocations		
	Prec A	Prec B	Prec C	Prec A	Prec B	Prec C
Block 2001	100	.	.	50	.	.
Block 2002	50	150	.	25	75	.
Block 2003	.	150	.	.	30	.
Block 2004	.	100	60	.	25	20
Block 2005	.	.	125	.	.	25

A simple example will illustrate this process. Suppose there are three precincts, A, B and C. There are five census blocks, 2001, 2002 2003, 2004 and 2005. Block 2001 is totally contained within precinct A, 2002 is split between A and B, 2003 is totally within B, 2004 is split between B and C, and 2005 is totally within precinct C. There are 225 registered voters in precinct A, 535 in B, and 230 in C. Of those registered voters in A, 150 are assigned through geocoding and 75 are unassigned, for B, 400 are assigned through geocoding and 135 are unassigned, and in C, 185 are assigned through geocoding and 45 are unassigned. The population of block 2001 is 200, that of 2002 is 400, that of 2003 is 240, that of 2004 is 280, and that of 2005 is 200.

The above optimization equations can then be written as:

$$\begin{aligned} t1 &= \frac{a_{A1} + u_{A1}}{p_1} \\ t2 &= \frac{a_{A2} + a_{B2} + u_{A2} + u_{B2}}{p_2} \\ t3 &= \frac{a_{B3} + u_{B2}}{p_3} \\ t4 &= \frac{a_{B4} + a_{C4} + u_{B4} + u_{C4}}{p_3} \\ t5 &= \frac{a_{C5} + u_{C5}}{p_5}, \end{aligned}$$

where the t_i are the ratios of block $200i$. Equating all of these gives a solution where the ratio of registration to population is for all blocks .75. The actual assignments are given in Table 1.

Converting Registration and Electoral Data to census geography

The counties have registration (RG) and Statement of Vote or consolidated (SOV) precincts as well as ballot groups for each election. Registration statistics are reported by RG precinct and election results are reported by SOV precinct and/or ballot group (ABSPREC). The counties also report the grouping of RG

precincts in SOV precinct and Ballot group (although the ballot group reporting is inconsistent and incomplete).

The counties also provide maps or GIS files showing registration precinct geography, but the geographic representations do not necessarily match the RG precincts for one or more of the following reasons:

1. Zero-voter precincts not included in tabular data.
2. County has precinct and subprecinct RG precincts and consolidations, but maps don't show subprecinct geography.
3. County has alpha-split precincts, separate RG precincts with common geography; voters separated alphabetically.
4. Unexplained inconsistency or incomplete maps from the counties.

We create map precincts (MPREC) to reflect the geography as consistently as possible. RR precincts are aggregations of RG precinct (tabular data) into MPRECs (geographic). (Generally speaking, Map Precincts are RR precincts.)

Because the resulting RR precincts may include RG precincts that are consolidated into different SOV precincts, we create a geographic consolidation known as SR precincts to contain whole RR and SOV precincts. These precinct conversions are used in database construction to estimate probabilities of voter propensities (see next section on "Disaggregation of Precinct Voting Results to Census Geography"). They can also be used by database users to perform analyses of racially polarized voting.

Disaggregation of Precinct Voting Results to Census Geography

The method of allocating (or breaking down, or disaggregating) votes to census geography is done by assigning an individual probability to each voter who voted in the election and aggregating these probabilities to the census geography (the block, at the lowest level), to obtain a total vote for the block. The is done with the following methodology.

- Partition voters into groups. There are technical reasons why these groups should be "homogeneous," which is a statistical concept for how well the voting behavior fits a statistical law (in this case the multinomial distribution–see below). In general, in California, we use Democrat/Republican/Independent as the groups, with provisions for the variation of minorities and economic status (see next item). If there are P precincts, then X_{gi} is the number of voters in group g, $g = 1, \ldots, G$ in precinct i, $i = 1, \ldots, P$.
- Estimate the overall voting probabilities in the state/district. The probability of a member of group g voting for a candidate in precinct i is denoted

by p_{gi}, and if v is the votes for that candidate, an equation similar to the following is used to estimated p_{gi}:

$$\text{argmin} \sum_{i=1}^{P} (v_i - \sum_g X_g p_{gi})^2,$$

where the argmin is taken over the p_{gi} (in actual fact a more complicated optimization is used but it will produce results close to this equation–at times a suitable estimate is used rather than performing this estimation, depending upon the contest and the difficulty of estimating).

There are too many p_{gi} to estimate each individual p_{gi}, so an average p_{gi} is estimated instead. The average used here is of the form $p_{gi} = p_g(z_i, \tau)$, where τ is a conformable vector to a set of characteristics z_i for precinct i (which is how one would include variations in minorities or economic status). Note that for a district contest, each estimation should be done for each district by itself (or suitably modified, with the use of dummy variables), whereas statewide contests can usually be estimated with all precincts.

- Adjust estimated probabilities to the precinct. The method of estimation described above does not ensure that the estimated precinct totals equal the actual precinct totals on a precinct by precinct basis (though the overall estimation is done so that the estimated district/state totals do match the overall district/state totals). The standard statistical methodology for adjusting estimations is followed, where a $\tilde{p}$ is estimated such that $v_i = \sum_g X_{gi} \tilde{p}_{gi}$. This $\tilde{p}_i$ (which is a G by 1 vector) is estimated by

$$\tilde{p}_i = \text{E}[\hat{p}_i | v_i] \approx p(z_i, \hat{\tau}) + \text{Cov}[K_i | v_i][\text{Var}[v_i]]^{-1}[v_i - \sum_g X_{gi} p(z_i, \hat{\tau})],$$

where the K_i is the response count of the groups for the i^{th} precinct, with the $(K_{1i}, \ldots, K_{Gi})$ being distributed multinomially (some distributional assumptions must be made in order to estimated the covariance of K_i with v_i).

These $\tilde{p}_i$'s allocate all of the votes for a candidate in each precinct to the individual voters in that precinct. Thus the sums of these votes by census block by all census blocks will equal the sum of the vote for the candidate. For a census block split between precincts (say precinct A and B), those voters in precinct A will have an assigned voting propensity of $\tilde{p}_A$ and those in precinct B will have an assigned voting propensity of $\tilde{p}_B$.

As an example, consider a two-precinct district as described in table 3 (in this table, quantities are suppressed for units in which they do not make sense). The two precincts in the district, A and B, each have a block wholly contained in the precinct (1001 for A and 1003 for B) and share a block, 1002. There are two groups in the electorate, X_1 and X_2. Through geocoding, it is known how

Table 3: Example of Disaggregation in a Two-precinct District

Precinct	v	$\hat{v}$	Block	X_1	$\hat{p}_1$	$\tilde{p}_1$	$\hat{v}_1$	X_2	$\hat{p}_2$	$\tilde{p}_2$	$\hat{v}_2$
A		150.0	1001	100		.2	20.0	200		.65	130.0
A		80.0	1002	50		.2	15.0	100		.65	65.0
A	230	230.0		150		.2	35.0	300		.65	195.0
B		56.5	1002	100		.34	34.0	50		.45	22.5
B		73.5	1003	150		.34	51.0	50		.45	22.5
B	130	130.0		250		.34	85.0	100		.45	45.0
District	360	360.0		400	.3		120.0	400	.6		240.0

many voters of each group is within each block and what precinct each voter is in, which is known through the registered voter rolls. The number of votes cast for a candidate (v) is known at the precinct level.

Estimated quantities are an overall probability of each group to vote for the candidate ($\hat{p}_1$ for group X_1 and $\hat{p}_2$ for group X_2) and then the adjusted probabilities ($\tilde{p}_1$ for group X_1 and $\tilde{p}_2$ for group X_2) chosen by the method described above in "Adjust estimated probabilities to the precinct". This then gives, for each precinct/block combination, an estimated vote by group ($\hat{v}_1$ for X_1 and $\hat{v}_2$ for X_2). Finally, the sum of $\hat{v}_1$ and $\hat{v}_2$ can be calculated, providing a $\hat{v}$ for each precinct/block combination and a check that, indeed, the allocated (or disaggregated) vote does equal the actual vote for each precinct.

For a block split between two (or more) precincts, these estimated totals can be added up across precinct/block combinations which contain that block to obtain totals for the block. For example, in block 1002 in Table I, there are 49 votes from group 1 for the candidate, 87.5 from group 2 for the candidate, for a total of 146.5 votes for the candidate.

Summary of P.L. 94-171 and CVAP Adjustments

For the 2020 Census, California is required to adjust the P.L. 94-171 data by allocating data from individuals incarcerated in state correctional facilities back to their last known residential address, as well as by excluding data from individuals incarcerated in federal correctional facilities. For purposes of this documentation, a "state correctional facility" means a facility under the control of the Department of Corrections and Rehabilitation. In addition, the Statewide Database (SWDB) has also been requested to allocate as best as possible the Citizen's Voting Age Population (CVAP) to the 2020 census geography. The processes used to create both datasets are discussed below.

P.L. 94-171 Adjustments Summary

Public Law 94-171 data consist of six "tables", with multiple entries for each table, including tabulations for total population, race, and ethnicity. These data are available at the level of the census block and can also be aggregated to larger units of analysis, such as the census tract. The California Department of Corrections and Rehabilitation (CDCR) has provided the SWDB with a list of persons incarcerated at state correctional facilities in the state of California and the last address of those persons prior to incarceration, if this address is available.

Election Code Section 21003 describes how data from persons enumerated at state correctional facilities are to be allocated to census geography at the level of the census block. The 2020 census enumerated these persons at the correctional facility in which they were incarcerated at the time of the scheduled census enumeration, and they are hence tallied in the census block of the correctional facility. Section 21003 specifies how these persons are to be removed from the census block associated with their place of incarceration and placed in the census block of their last known place of residence. This is done using the census address/block conversion system (the TIGER files) when a sufficiently complete address is available. When a sufficiently complete address is not available, this is done through random assignment to a census block that is part of the most restrictive level of census geography that can be determined from the incomplete address.

Adjustments are also made to account for persons who reside in census blocks that are associated with state correctional facilities but who are not themselves incarcerated, and also for persons in federal custody in facilities in California (this is described in the section "P.L. 94-171 Adjustments"). The final total population counts (pre and post adjustments) are as follows:

Table 4: Population counts for California

Source	Population
US Census P.L. 94-171	39,538,223
Adjusted P.L. 94-171	39,523,437

CVAP adjustments summary

The Citizen's Voting Age Population (CVAP) special tabulation derives from a census product known as the American Community Survey (ACS), located at URL https://www.census.gov/programs-surveys/acs. These surveys ask for information on the citizenship of the respondent, albeit for a different time period than the P.L. 94-171 enumeration (CVAP data derive from survey responses over the period 2015-2019 as compared to one day for the P.L. 94-171). The surveys are also a subset of the population, being a survey, while the P.L. 94-171

attempts a complete enumeration of the population on a particular date.

Aside from the incompatibility of time periods and collection periods, the survey data are reported by units of the 2010 census geography whereas the P.L. 94-171 data are reported by units of the 2020 census geography. While the Census Bureau provides a conversion table, geographic conversions are typically inaccurate. Finally, as is demonstrated in the body of this documentation, even when errors cannot be ascribed to difficulties of geographical conversion, there are still fairly significant discrepancies between the CVAP figures and the P.L. 94-171 figures (see section "Creating 2010 Block Group to 2020 Block Conversions").

Similar to the P.L. 94-171 adjustment, CVAP adjustments are made for incarcerated persons. The final CVAP total citizen population counts in California (as well as P.L. 94-171 counts) are given in table 5 The following sections ("Data

Table 5: P.L. 94-171 and ACS Population counts

Source	Population
P.L. 94-171 Population	39,538,223
Adjusted P.L. 94-171 Population	39,523,437
ACS Population Estimate	39,283,495
ACS Citizen Population Estimate	34,187,375
Adjusted ACS Citizen Population Estimate	34,188,096

from Incarcerated Persons and P.L. 94-171 Adjustments" and "CVAP Data") explain how these numbers are obtained.

Data from Incarcerated Persons and P.L. 94-171 Adjustments

Input Data

A list of incarcerated persons was obtained from the California Department of Corrections and Rehabilitation (CDCR), this list being current on or about April 1, 2020. The data fields in this list were as follows: There are 122,730 unique ids in this list, from a total of 159,111 records. Ethnicities were provided and were recoded to approximately fit census categories, as indicated in table 7. It is assumed that all the incarcerated individuals in state facilities are 18 or older. Thus the same modifications made to table P2 are made in the same manner to table P4. Table 8 has the counts for the ethnicities of incarcerated persons, recoded to fit the census P.L. 94-171 categories.

Geocoding Procedure

All incarcerated persons in state facilities (represented by unique ids) were to be placed in the most restrictive geography available. Ideally, each incarcerated

Table 6: Data fields of incarcerated person data

Unique ID
GQ Name
ETHNICITY
Race
Address Number
Street Name
Apt/Unit
Rural Route Address
City
State
ZipCode
Effective Date
Location Description
Lifer
County Of Last Legal Residence

person would have a sufficient address to be able to be placed in a census block using the Census Bureau's address block equivalency files (the TIGER files). As an (idealized) example, such a conversion file would have entries of the form: Thus an address of 124 N Elm St would be placed in census block 061170001002001 (the census block format is given in table 10) Addresses are unique within counties so a "complete" address would have an address number, a street name, and a county (street type and direction can also be specified–note that city and/or zip is not necessary for a complete address). The success rate of this geocoding is given in table 11.

Procedure for addresses unable to be geocoded

The criteria for obtaining the most restrictive geographical area were implemented as follows in the following order for the 53,688 incarcerated persons who were not geocoded. The 59 who were geocoded to their place of incarceration were assigned to the State Match pool (see below).

- Extending the Address:

  If an address was "complete" (street number, street name, county), but the street number did not have a block conversion (but the street name did and the county could be ascertained), then an attempt was made to match the address by finding the block closest to the street number. This was done in the following two steps:

  - Address Match County

    If an address had an address number, street name and county, a match was attempted on the closest TIGER conversion file entry of

Table 7: Counts by ethnicity of incarcerated persons with recoded census category

ethnicity	count(*)	recoded
Filipino	412	asian
White	25322	white
Other	4828	other
Hispanic	38918	latino
Black	34718	black
Mexican	15428	latino
American Indian	1403	amIndian
Vietnamese	239	asian
Pacific Islander	163	pacific
Salvadorian	99	latino
Laotian	56	asian
Korean	57	asian
Other Asian	368	asian
Puerto Rican	98	latino
Thai	18	asian
Chinese	104	asian
Cuban	57	latino
Hawaiian	40	pacific
Japanese	25	asian
Samoan	148	pacific
Cambodian	89	asian
Guatemalan	61	latino
Guamanian	22	pacific
Indian	33	asian
Jamaican	4	black
Columbian	6	latino
Nicaraguan	11	latino
Unknown	3	other

Table 8: Ethnicity recoding for compatibility with P.L. 94-171

ethnicity	number	P2 variable	P4 variable
latino	54678	P0020002	P0040002
black	34722	P0020006	P0040006
white	25322	P0020005	P0040005
other	4831	P0020010	P0040010
amIndian	1403	P0020007	P0040007
asian	1401	P0020008	P0040008
pacific	373	P0020009	P0040009

Table 9: Example of idealized address to census block conversion

start	stop	street	type	direction	parity	block
100	198	Elm	St	N	even	061170001002001

Table 10: Census block code format

digits	field description
1-2	State (06 for California)
3-5	County (FIPS code–Federal Information Processing System code)
6-11	Tract
12-12	Block Group
12-15	Block

address to block. This could fail if the street was not in the TIGER conversion file.

If this was not successful, then the following was tried:

– Address Match City

If an address had an address number, street name and no county, but did have a city, the city was placed in a county and the procedure under "Address Match County" was followed. Once again, the street needed to be valid.

- City Match:

If the proceeding two steps under Extending the Address failed, and if the city was present and valid, assignment was made to a randomly selected block within the city. This randomization was proportional to the population of the block–that is, if block A has twice the population of block B, then block A has twice the probability of being selected as the block of the incarcerated person as block B. Formulaically,

$$\Pr\left[\text{block A being chosen}\right] = \frac{\text{Population of block A}}{\sum \text{Population of all blocks in city}}$$

- County Match:

Table 11: Geocoding status of addresses for incarcerated persons

69,042	successfully geocoded
53,688	not geocoded
122,730	total
59	geocoded to place of incarceration

Table 12: Incarcerated person assignments by succeeding steps

Specification					Address Match		City	County
Street	City	Zip	County	Initial	County	City	Match	Match
N	N	N	N	6034	6034	6034	6034	6034
N	N	N	Y	32342	32342	32342	32342	122
N	Y	N	N	20	20	20	20	20
N	Y	N	Y	19	19	19	15	0
N	Y	Y	N	3	3	3	2	2
N	Y	Y	Y	77	77	77	14	0
Y	Y	N	N	399	399	399	366	366
Y	Y	N	Y	1539	1172	1167	601	36
Y	Y	Y	N	859	859	859	459	459
Y	Y	Y	Y	12396	7806	7737	603	6
*A field being specified does not mean that it is valid								

If the proceeding steps failed to make an assignment, an assignment to a county was attempted if the county was specified. The randomization method is the same as with city, but here the probability of a block in the county was calculated using the population of the entire county.

$$\Pr\left[\text{block A being chosen}\right] = \frac{\text{Population of block A}}{\sum \text{Population of all blocks in county}}$$

- State Match:

  If the proceeding steps failed to make an assignment, an assignment was made to all blocks in the state. The randomization method is the same as with city or county, but here the probability of a block in the state was calculated using the population of the entire state.

  $$\Pr\left[\text{block A being chosen}\right] = \frac{\text{Population of block A}}{\sum \text{Population of all blocks in state}}$$

  The 59 incarcerated persons geocoded to the census block of their incarceration were allocated by this method.

It should be noted that city, county, and state match are randomly assigning the address to the most restrictive geography.

Incarcerated person assignments remaining to be made after each succeeding step of the algorithm, by input address specification:

In this table, the initial column provides the breakdown of the 53,688 who could not be geocoded. For the first step (Extending the Address: Address Match County), 4,597 are assigned [(12396-7806) + (1539 - 1172)]. Relatively few are

Table 13: Geographical assignment type counts for incarcerated persons

type	number
geocoded	68983
street	5031
city	8183
county	33411
state	7122

assigned using the city in place of the county (74 in Extending the Address: Address Match City). Note the zip code field is in the specification but corrections based on the zip code substitution are not made part of the algorithmic process, as there were so few specified and the Address Match City returned so few matches. Finally, the City Match part of the algorithm has thousands of matches and the County Match has tens of thousands. Any unassigned incarcerated persons after the County Match (and the 59 with correctional facility geocoded addresses) are handled by the State Match part of the algorithm, so that all incarcerated persons are placed in a census block. The 122,730 incarcerated persons are assigned to 81,648 blocks.

The final totals for all the categories of assignments are given in table 13 (in this table, street combines Address Match County and Address Match City): The state pool, having the least geographic specificity, was the pool chosen when incarcerated persons were removed due to the CDCR reporting more incarcerated persons than were enumerated by the census (see section "P.L. 94-171 Adjustments").

Allocation

Data on incarcerated persons obtained from the California Department of Corrections and Rehabilitation (CDCR) included an address (with varying levels of completeness) and an ethnicity. The P.L. 94-171 census data for population come in five tables. Tables P1 and P3 are organized by race and are not suited for updating with incarcerated populations, who are classified in the CDCR data with a combined race/ethnicity variable. Tables P2 and P4 are of a form that lend themselves to a straight-forward conversion from the CDCR race/ethnicity variable and these are the tables that are updated. Table P5 deals with population in group quarters and while this table is not adjusted, it is used in the adjustment process.

The census geography from which incarcerated persons are to be reallocated is as follows.

- Census blocks were designated as state correctional facility blocks, primarily on the basis of actual state correctional facility locations but with some additions of blocks that had incarcerated individuals who were not in

Table 14: Assignment of incarcerated persons to census block type

All incarcerated persons		Incarcerated persons minus deleted		
type	number	type	number	Difference
geocoded	68983	geocoded	68983	0
street	5031	street	5031	0
city	8183	city	8183	0
county	33411	county	33411	0
state	7122	state	6785	337

county facilities (P.L. 94-171 Table 5, variable 3 (P0050003)). Thus each correctional facility had one or more census blocks associated with it (see section "Correctional Facility Census Block Adjustment" for the blocks modified).

- If there was a greater population (P.L. 94-171 Table P2, variable 1 (P0020001) than group quarters (P0050003) the population of that block was set to P0020001 minus P0050003.
- Allocations of ethnicity for this remainder population was done on the basis of the underlying census categories for the entire census block.
- There were 337 additional incarcerated persons in the CDCR list compared with the census enumeration, defined as the total from the CDCR list (122,730) minus the group quarters variable P0050003 (122,393). These additional incarcerated persons were randomly removed from the type referred to as "state" in the below table (and described under "State Match" in section "Allocation of Incarcerated Persons"), as this "state" assignment was randomly made (exact procedures for the allocation of incarcerated persons to the most restrictive geographic area are described in section "Allocation of Incarcerated Persons").
- Due to the lack of data on last known residential address, individuals in federal custody who lived in group quarters (P0050003) were removed from the adjusted P.L. 94-171 data. The non-group quarters population in the federal facility blocks (see section "Correctional Facility Census Block Adjustment") was kept in the adjusted P.L. 94-171 (there were 67 such people).
- The adjusted population total for California is thus equal to

  P0020001 - Persons in Federal Custody

  and the additions and subtractions to obtain this figure are as follows:
- There were two versions of the adjusted P.L. 94-171 released by the SWDB– one on September 20 and another on September 27. There were two differences–first, adjustments were made to tables 2 and 4 of the adjusted

Table 15: Calculation of final P.L. 94-171 population figures

39,538,223	Original P.L. 94-171 Count for P0020001
122,393	Group Quarters in state correctional facility blocks (removed from P.L. 94-171)
39,415,830	P.L. 94-171 Count after group quarters removed
122,730	Number of incarcerated persons
122,393	Group Quarters in state correctional facility blocks
337	Additional incarcerated persons fron CDCR list (to be removed randomly from the allocated in
39,538,223	P.L. 94-171 Count with incarcerated persons added
14,786	Group Quarters in federal correctional facility blocks (removed from P.L. 94-171)
39,523,437	Final P.L. 94-171 count

Table 16: Incarcerated person ethnicity breakdown before and after deletions

All incarcerated persons		Incarcerated persons minus deleted		
Classification	Number	Classification	Number	Difference
latino	54,678	latino	54,527	151
black	34,722	black	34,638	84
white	25,322	white	25,270	52
other	4,831	other	4,795	36
amIndian	1,403	amIndian	1,396	7
asian	1,401	asian	1,395	6
pacific	373	pacific	372	1
total	122,730	total	122,393	337

data so the equated quantities in the unadjusted P.L. 94-171 for these tables were maintained in the adjusted data (for example, P0020001 = P0020002 + P0020003). Second, the random selection of additional incarcerated persons, to account for the CDCR reporting 337 more incarcerated persons than the census, was a different random selection than the previous. These additional incarcerated persons were selected from the randomized assignment pool (the state match pool–see section "Allocation of Incarcerated Persons").

- The race/ethnicity breakdowns from the incarcerated person data (before and after deletions) are given in the table below. This is a condensation of the reported categories from the CDCR to fit tables P2 and P4 from the P.L. 94-171 data (the condensation is described in section "Ethnicity of Incarcerated Persons").

Correctional Facility Census Block Adjustment

The procedure for allocating data about incarcerated persons to their last known residential address requires identifying census blocks that have adults incarcerated in correctional facilities (P.L. 94-171 Table P5, variable number 3 – P0050003). All blocks with non-zero P0050003 are included in the allocation procedure. The correspondence between the physical location of correctional facilities as reported by the CDCR and the location of non-zero P0050003 census blocks is not exact, but in general there is a relatively close geographic agreement between them. Some non-zero P0050003 census blocks were reported at a greater distance, however, making locating and matching them a nontrivial task. The count of incarcerated persons reported by the CDCR likewise did not exactly match the count reported in the Census P.L. 94-171 Table P5. The reasons for this are likely due to the Census Bureau's new privacy methodologies, in particular the use of Differential Privacy, in addition to Census Bureau errors in geocoding and with group quarters data reporting in general. Where this was the case, the reallocation procedure relied on the reported data in Table 5 of the P.L. 94-171 to make the adjustments.

Tables 17, 18, and 19 list census blocks associated with state correctional facilities. There are two facilities that have a "null" block of 000000000000000: PRMCCF-Golden State and the "Legal Processing Unit". These are facilities which may have incarcerated persons in the list provided by the CDCR but do not have specified physical locations (PRMCCF-Golden State is the Female Community ReEntry Facility).

CVAP Data

The latest CVAP (Citizen Voting Age Population) data available from the Census Bureau were from the 2015-2019 ACS. These data are reported by the block group from 2010 census geography. There are 14 data "lines" for each block group (displayed in table 21.

A "line" is the Census Bureau's description of the four data, displayed in table 22. The two data that are of primary interest are the cit_est and the cvap_est.

In the ACS there are 23,212 block groups. These block groups are in 2010 geography. To convert between 2010 and 2020 geography, a conversion file provided by the Census Bureau is used. Certain block groups in the 2010 ACS exist under different designations than in this 2010 to 2020 conversion file. A conversion is made between these designations for the 2010 ACS and then the converted block designations are treated the same as all other blocks in the 2010 data.

where the replacement block group is in the 2010 to 2020 conversion file while the replaced block group is what is in the ACS. A full list of 2010 replaced blocks in the ACS is below. Once these blocks are renamed, they can then be used as

Table 17: Census blocks adjusted due to state correctional facilities-1

Census block	Prison
060319818001000	Avenal State Prison
060855120052006	Avenal State Prison
060290065002618	California City Correctional Facility
060372035001009	California City Correctional Facility
060372971101013	California City Correctional Facility
060350404002006	California Correctional Center
060290060021000	California Correctional Institution
060014340003003	California Correctional Institution
060770051311007	California Health Care Facility - Stockton
060710005041028	California Institution for Men
060710122021007	California Institution for Men
060710122012012	California Institution for Men
060710122021003	California Institution for Men
060710019071004	California Institution for Women
060710098001017	California Institution for Women
060952530001001	California Medical Facility
060170312001028	California Medical Facility
060290016002083	California Medical Facility
060790114001000	California Men's Colony
060372312201007	California Men's Colony
060650466011001	California Rehabilitation Center
060375307001034	California Rehabilitation Center
060319801001002	California State Prison, Corcoran
060379010031002	California State Prison, Los Angeles County
060374822011009	California State Prison, Los Angeles County
060679883001003	California State Prison, Sacramento
060375780004010	Folsom State Prison
060952530001007	California State Prison, Solano
060730100053003	California State Prison, Solano
060319801001004	California Substance Abuse Treatment Facility
060250101011274	Calipatria State Prison
060250101011115	Calipatria State Prison
060375331083002	Calipatria State Prison
060150002031222	CCC-Alder Camp
060350404002002	CCC-Antelope Camp
060871202002042	CCC-Ben Lomond Camp
060871224021016	CCC-Ben Lomond Camp
060450103005066	CCC-Chamberlain Creek Camp
060930007012059	CCC-Deadwood Camp
060952535013187	CCC-Delta Camp

Table 18: Census blocks adjusted due to state correctional facilities-2

Census block	Prison
060490003001326	CCC-Devils Garden Camp
060230115012058	CCC-Eel River Camp
060330006023071	CCC-Eel River Camp
060230111003084	CCC-High Rock Camp
060350401001126	CCC-Intermountain
060372324021001	CCC-Intermountain
061030001001291	CCC-Ishi Camp
060710121014108	CCC-Ishi Camp
060330009012016	CCC-Konocti Camp
060450103005097	CCC-Parlin Fork Camp
061030003003220	CCC-Salt Creek Camp
060890126061146	CCC-Sugar Pine Camp
061050001021118	CCC-Trinity Camp
060375990002007	CCC-Trinity Camp
060030100001048	CCC-Trinity Camp
060210103001010	CCC-Valley View Camp
060570008012037	CCC-Washington Ridge Camp
060250123012161	Centinela State Prison
060390002011025	Central California Women's Facility
060371133012014	Central California Women's Facility
060390002011024	Central California Women's Facility
060659810001001	Chuckawalla Valley State Prison
060659810001003	Ironwood State Prison
060659810001002	Ironwood State Prison
060378004061009	CIW-Malibu Camp
060375436011003	CIW-Malibu Camp
060730209033065	CIW-Puerta La Cruz
060730190021001	CIW-Rainbow Camp
060371959031002	Community Prisoner Mother Program
060530109001002	Correctional Training Facility
060372347004000	Correctional Training Facility
060371064071000	Correctional Training Facility
060375780004011	Correctional Training Facility
061110075131001	Correctional Training Facility
060372088012000	Correctional Training Facility
060530109001006	Correctional Training Facility
060779800001000	Deuel Vocational Institution
060770031161004	Deuel Vocational Institution
060290047032003	Female Community ReEntry Facility
000000000000000	PRMCCF-Golden State
060350404002001	High Desert State Prison

Table 19: Census blocks adjusted due to state correctional facilities-3

Census block	Prison
060290046011001	Kern Valley State Prison
060290046031000	North Kern State Prison
060290019023020	North Kern State Prison
000000000000000	Legal Processing Unit
060050003012002	Mule Creek State Prison
060050003012007	Mule Creek State Prison
060150002013066	Pelican Bay State Prison
060190079032083	Pleasant Valley State Prison
060290050051016	PUMCCF-Delano
060290040011038	PUMCCF-Shafter
060290033043168	PUMCCF-Taft
060730100161015	RJ Donovan Correctional Facility
060730123021013	RJ Donovan Correctional Facility
060670020001001	Sacramento Control Office
060670052051025	Sacramento Control Office
060670006001011	Sacramento Control Office
060670053011046	Sacramento Control Office
060530109001001	Salinas Valley State Prison
060411212002004	San Quentin State Prison
060411220001007	San Quentin State Prison
060379108153144	SCC-Acton Camp
061090052011010	SCC-Baseline Camp
060650444071025	SCC-Bautista Camp
060379304003047	SCC-Fenner Camp
060379200501181	SCC-Francisquito Camp
060530109001015	SCC-Gabilan Camp
060170306062038	SCC-Growlersburg Camp
060379304002183	SCC-Holton Camp
060379304003187	SCC-Julius Klein Camp
060730209021044	SCC-La Cima Camp
060730209021092	SCC-La Cima Camp
060730211021027	SCC-McCain Valley Camp
060190064102039	SCC-Miramonte Camp
061070027011022	SCC-Mountain Home Camp
060430001014058	SCC-Mt. Bullion Camp
060710115001114	SCC-Oak Glen Camp
060270002001117	SCC-Owens Valley Camp
060710108061000	SCC-Pilot Rock Camp
060710122021006	SCC-Prado Camp
060090001223038	SCC-Vallecito Camp
061110093002111	SCC-Ventura Conservation Camp
061070020091000	SHS-Atascadero State Hospital
060190011002042	SHS-Coalinga State Hospital
060710041032019	SHS-Patton State Hospital
061099852021000	Sierra Conservation Center
060390002011026	Valley State Prison
060290043021001	Wasco State Prison
060290005043008	Wasco State Prison
060371902012004	Wasco State Prison
060371895022000	High Desert State Prison
060375042002035	SCC-Holton Camp

Table 20: Census blocks adjusted due to federal correctional facilities

Census block	Prison
060014501021046	FCI Dublin
060190083041101	FCI Mendota
060350406002032	FCI Herlong
060372074001029	MDC Los Angeles
060379800311015	FCI Terminal Island
060470005031174	USP Atwater
060719802001033	FCI Victorville
060730053021025	MCC San Diego
060730053021030	MCC San Diego
060839804001008	FCI Lompoc
060839804001011	USP Lompoc
060290033043171	

Table 21: ACS/CVAP ethnic breakdowns

Line	Description
1	Total
2	Not Hispanic or Latino:
3	American Indian or Alaska Native Alone
4	Asian Alone
5	Black or African American Alone
6	Native Hawaiian or Other Pacific Islander Alone
7	White Alone
8	American Indian or Alaska Native and White
9	Asian and White
10	Black or African American and White
11	American Indian or Alaska Native and Black or African American
12	Remainder of Two or More Race Responses
13	Hispanic or Latino

Table 22: ACS/CVAP data fields

Field name	Description
CIT_EST	The rounded estimate of the total number of United States citizens for that geographic area and group
CIT_MOE	The margin of error for the total number of United States citizens for that geographic area and group
CVAP_EST	The rounded estimate of the total number of United States citizens 18 years of age or older for that geographic area and group
CVAP_MOE	The margin of error for the total number of United States citizens 18 years of age or older for that geographic area and group

Table 23: Block Groups: 2010/2020 Block Conversion vs. 2010 ACS Geography

replacement_bg	replaced_bg	nblks
060378002043	060371370002	29
060379304011	060371370001	25
060379304011	060371370002	5

other 2010 blocks are used in the breakdown procedure.

Rather than refer to the CVAP line numbers, we use the names given in table 25

CVAP adjustments

To adjust the CVAP data to the P.L. 94-171, it is necessary to create a conversion between the P.L. 94-171 tables and the CVAP lines. The relevant P.L. 94-171 variables are given in table 26.

The equivalencies between the CVAP data and the P.L. 94-171 data are given in the following table:

Table P2 is used because the CVAP data splits its sample between Hispanic or Latino and Not Hispanic or Latino, as does table P2.

A similar comparison can be made for citizen voting age population and over 18 by substituting in the appropriate tables (table P3 for P1 and table P4 for P2 for the P.L. 94-171, and cvap_est for cit_est for the ACS). These categories are used for breaking down the ACS data to the census block by first estimating an overall proportion for the entire block group for each of the CVAP variables (26 in total: 13 for citizen, 13 for citizen voting age) to the equivalent block group variables or combination of variables in the P.L. 94-171 data, and then using this estimated proportion for each census block within the block group, or, formulaically,

$$\text{estimated\_ratio} = \text{cvap variable/pl94 variable at block group level} \quad (1)$$

$$\text{block\_cvap\_variable} = \text{estimated\_ratio} * \text{pl94\_block\_variable} \quad (2)$$

This procedure usually results in a close allocation of the CVAP to the 2020 census block geography. The closeness metric is defined with the following algorithm:

- Obtain the original CVAP population (line 1) at the 2010 block group level
- Create 2010 block group to 2020 block conversions from census files (geographic)

Table 24: 2010 Blocks Replaced in the ACS

replacement_bg	replaced_bg
060379304011000	060371370001001
060379304011001	060371370001002
060379304011002	060371370001003
060379304011003	060371370001004
060379304011004	060371370001005
060379304011005	060371370001006
060379304011007	060371370001007
060379304011008	060371370001008
060379304011009	060371370001009
060379304011010	060371370001010
060379304011011	060371370001011
060379304011012	060371370001012
060379304011016	060371370001013
060379304011017	060371370001014
060379304011019	060371370001015
060379304011020	060371370001016
060379304011021	060371370001017
060379304011022	060371370001018
060379304011023	060371370001019
060379304011024	060371370001020
060379304011025	060371370001021
060379304011026	060371370001022
060379304011027	060371370001023
060379304011028	060371370001024
060379304011029	060371370001025
060378002043001	060371370002001
060378002043002	060371370002002
060378002043003	060371370002003
060378002043004	060371370002004
060378002043005	060371370002005
060378002043006	060371370002006
060378002043007	060371370002007
060378002043008	060371370002008
060378002043009	060371370002009
060378002043010	060371370002010
060378002043011	060371370002011
060378002043012	060371370002012
060378002043013	060371370002013
060378002043014	060371370002014
060378002043015	060371370002015
060378002043016	060371370002016
060378002043017	060371370002017
060378002043018	060371370002018
060378002043019	060371370002019
060378002043020	060371370002020
060378002043021	060371370002021
060378002043022	060371370002022
060378002043023	060371370002023
060378002043024	060371370002024
060378002043025	060371370002025
060378002043026	060371370002026
060378002043027	060371370002027
060378002043028	060371370002028

Table 25: ACS/CVAP variable names

Line field	Variable name
line1.CIT_EST	ctot
line2.CIT_EST	cnonhisp
line3.CIT_EST	camIndian
line4.CIT_EST	casian
line5.CIT_EST	cblack
line6.CIT_EST	cpacific
line7.CIT_EST	cwhite
line8.CIT_EST	camIndian_white
line9.CIT_EST	casian_white
line10.CIT_EST	cblack_white
line11.CIT_EST	camIndian_black
line12.CIT_EST	cother_two
line13.CIT_EST	clatino
line1.CVAP_EST	cvtot
line2.CVAP_EST	cvnonhisp
line3.CVAP_EST	cvamIndian
line4.CVAP_EST	cvasian
line5.CVAP_EST	cvblack
line6.CVAP_EST	cvpacific
line7.CVAP_EST	cvwhite
line8.CVAP_EST	cvamIndian_white
line9.CVAP_EST	cvasian_white
line10.CVAP_EST	cvblack_white
line11.CVAP_EST	cvamIndian_black
line12.CVAP_EST	cvother_two
line13.CVAP_EST	cvlatino

Table 26: Relevant P.L. 94-171 variables

P.L. 94-171	P.L. 94-171 Description
P0020001	Total
P0020003	Not Hispanic or Latino:
P0020007	American Indian and Alaska Native alone
P0020008	Asian alone
P0020006	Black or African American alone
P0020009	Native Hawaiian and Other Pacific Islander alone
P0020005	White alone
P0020014	White; American Indian and Alaska Native
P0020015	White; Asian
P0020013	White; Black or African American
P0020018	Black or African American; American Indian and Alaska Native
P0020002	Hispanic or Latino
P0020012	Population of two races
P0020014	White; American Indian and Alaska Native
P0020015	White; Asian
P0020013	White; Black or African American
P0020018	Black or African American; American Indian and Alaska Native
P0020028	Population of three races
P0020049	Population of four races
P0020065	Population of five races
P0020072	Population of six races

Table 27: CVAP and P.L. 94-171 equivalencies

ACS variable	P.L. 94-171 conversion
ctot	P0010001
cnonhisp	P0020003
camIndian	P0020007
casian	P0020008
cblack	P0020006
cpacific	P0020009
cwhite	P0020005
camIndian_white	P0020014
casian_white	P0020015
cblack_white	P0020013
camIndian_black	P0020018
clatino	P0020002
cother_two	P0020012
	- P0020014
	- P0020015
	- P0020013
	- P0020018
	+ P0020028
	+ P0020049
	+ P0020065
	+ P0020072

Table 28: Comparison of original ACS population and reassembled by 2010 block groups

difference	number
-5 <, < 5	22003
> 5	623
<-5	586

- Break down original CVAP population (step 1) at the 2010 block group level to the 2020 census block level, using the census conversions (step 2)
- Reassemble these data by summing the broken down CVAP 2020 census block data (step 3) to the 2010 census block group, using the census conversions (step 2)
- Compare original CVAP at the 2010 block group level (step 1) to the reassembled data (step 4)

Breaking down the data and reassembling them in this manner to make a comparison creates error in both steps. However, the accuracy of the conversion appears to be good: the difference in allocation between these two quantities is essentially equivalent in about 95% of the cases, as displayed in table 28.

The more extreme differences (the absolute value of the difference being more than five) come about in cases where 2020 blocks are split geographically between multiple 2010 block groups, and the geographic assignments are suspect. These differences (which are a relatively small in number–a few thousand in a 35 million or so citizen total population) are handled by assigning differences to geography within the county, so that county totals for CVAP data by 2010 block groups and 2020 blocks agree at this stage of the allocation process. The next step will be to reallocate data from persons who were incarcerated in state correctional facilities at the time of the census to their last known residential address. When that is done, there will no longer be an agreement between the county totals for the CVAP on 2010 census geography and the CVAP allocated to the 2020 census geography.

The above discusses agreement of the data allocated from the CVAP by the 2010 census block group geography to the 2020 census block geography. There is another type of agreement, which we will term constraints, that some census products contain. For example, P.L. 94-171 data typically obeys certain constraints within a block group, that is, a set of variables summed equals another variable (for example, total population is equal to the sum of a certain number of P.L. 94-171 data variables in table P1).

The CVAP data, however, does not in general honor any particular set of

constraints. For example, the following equation should always hold:

$$\begin{aligned} 0 &= \text{ctot} - (\text{camIndian} + \text{casian} + \text{cblack} + \text{cpacific} + \text{cwhite} + \text{camIndian\_white} \\ &+ \text{casian\_white} + \text{cblack\_white} + \text{camIndian\_black} + \text{cother\_two} + \text{clatino}) \end{aligned} \quad (3)$$

But in actual fact the distribution of the right-hand side of (3) among the CVAP 2010 block groups is often different, as shown in table 29.

A comparison of (3) with the CVAP allocated to the 2020 census geography (summed up to the block groups for comparison with the above) shows a close equivalency, as displayed in 30. So about 87% of the 2020 block groups are within plus or minus 20 people of the constraint. Thus the constraint bounds are somewhat looser with the CVAP data allocated to the 2020 census blocks compared to the unallocated data, but still comparable.

CVAP for Incarcerated Persons

Since the P.L. 94-171 data are adjusted for the correctional facilities, the CVAP data also must be adjusted (see the "Allocation" section of "Data from Incarcerated Persons and P.L. 94-171 Adjustments"). There are 130 state correctional facility blocks (see tables 17, 18, and 19). All incarcerated persons were treated as citizens and were allocated to their assigned geocoded blocks, their ethnicity being converted to CVAP variables by the calculus in table 31. 111,820 incarcerated persons were allocated using this rule. Each incarcerated person was treated as a citizen and as 18 and over.

Adjustments to the census blocks that were associated with correctional facilities were also made. The vast majority (118,820) of the population of the correctional facility census blocks (122,730) were reallocated using this method. The citizenship data for the remainder of the population was handled following the algorithm set out in P.L. 94-171 adjustments. If there was a greater population (P.L. 94-171 Table 2, variable 1 (P0020001) than group quarters (P0050003) the population of that census block was set to P0020001 minus P0050003, and citizen data was assigned in the following manner. First, if P0020001 minus P0050003 were zero, no citizen data was assigned to that block. If this quantity were positive (it can never be negative), multiple normalizations and adjustments were made (there were 55 such blocks).

The algorithm that was then used for allocation of ACS citizenship data to census blocks was as follows (CIT is citizenship, CVAP is Citizen Voting Age Population):

- Calculated percent remaining based on P.L. 94-171 (P0010001 - P0050003)/P0010001
- Multiplied all CIT, CVAP variables by that percentage, giving a total citizenship number of 11,116 non-incarcerated persons in those blocks, and 111,820 incarcerated persons (total CIT was 122,936 in the correctional facility blocks)

Table 29: Differences of ACS total population from subcategories

diff	n
-17	1
-16	1
-15	1
-14	4
-13	5
-12	3
-11	2
-10	158
-9	121
-8	94
-7	31
-6	7
-5	3748
-4	1036
-3	318
-2	75
-1	11
0	10401
1	2136
2	414
3	75
4	12
5	3343
6	811
7	181
8	27
9	3
10	118
11	58
12	11
13	3
14	1
15	2

Table 30: Differences of allocated 2020 ACS total population from sub categories

diff	n block groups
< -20	1447
> 20	1972
$[-20, 20]*$	22188
*[] is the closed interval	

Table 31: Translation of incarcerated person data to CVAP data

ethnicity condition	Operation	CVAP variable
incarcerated_persons.ethnicity = any	add 1	bctot
incarcerated_persons.ethnicity != "latino"	add 1	bcnonhisp
incarcerated_persons.ethnicity = "amIndian"	add 1	bcamIndian
incarcerated_persons.ethnicity = "asian"	add 1	bcasian
incarcerated_persons.ethnicity = "black"	add 1	bcblack
incarcerated_persons.ethnicity = "pacific"	add 1	bcpacific
incarcerated_persons.ethnicity = "white"	add 1	bcwhite
incarcerated_persons.ethnicity = any	add 0	bcamIndian_white
incarcerated_persons.ethnicity = any	add 0	bcasian_white
incarcerated_persons.ethnicity = any	add 0	bcblack_white
incarcerated_persons.ethnicity = any	add 0	bcamIndian_black
incarcerated_persons.ethnicity = "other"	add 1	bcother_two
incarcerated_persons.ethnicity = "latino"	add 1	bclatino
incarcerated_persons.ethnicity = any	add 1	bcvtot
incarcerated_persons.ethnicity != "latino"	add 1	bcvnonhisp
incarcerated_persons.ethnicity = "amIndian"	add 1	bcvamIndian
incarcerated_persons.ethnicity = "asian"	add 1	bcvasian
incarcerated_persons.ethnicity = "black"	add 1	bcvblack
incarcerated_persons.ethnicity = "pacific"	add 1	bcvpacific
incarcerated_persons.ethnicity = "white"	add 1	bcvwhite
incarcerated_persons.ethnicity = any	add 0	bcvamIndian_white
incarcerated_persons.ethnicity = any	add 0	bcvasian_white
incarcerated_persons.ethnicity = any	add 0	bcvblack_white
incarcerated_persons.ethnicity = any	add 0	bcvamIndian_black
incarcerated_persons.ethnicity = "other"	add 1	bcvother_two
incarcerated_persons.ethnicity = "latino"	add 1	bcvlatino

- Allocated the 111,820 incarcerated persons to their geocoded blocks, not allocating 10,910 incarcerated persons. Incarcerated persons were removed on the basis of their assignment category, with state assignment types being removed completely and then county assignment types to make up the difference.

The earlier September 20th release utilized random matching on the L1 norm of the difference between the P2 variables (1 through 10) for census blocks associated with the correctional facilities and all other blocks, choosing the block that minimized that norm. This methodology produced some counter-intuitive results for a few blocks. The September 27th release utilized the final methodology described in this section, which produced fewer anomalous results. Some counter-intuitive results are to be expected with any adjustment methodology for these blocks, because two separate data sources (CVAP and P.L. 94-171) are being combined to produce the adjustment.

Creating 2010 block group to 2020 block conversions

The conversion file between the 2010 census geography for the CVAP data and the 2020 census geography is constructed from the Census Bureau conversion between 2010 and 2020 census blocks. The conversion between blocks need not be exact–that is, there can be multiple 2010 blocks in a 2020 block and vice versa.

A conversion file for a block group from the 2010 geography is simply the aggregation of all of the conversions from the 2010 blocks that constitute a block group to the 2020 blocks that are associated with the 2010 blocks (note that a block group GEOID is simply the block GEOID without the last three digits specified–see table 10).

An example of how conversions need not be exact is given in table 32. This table shows block group conversions for county 121 (Yuba is the highest county in California with a FIPS code of 115, so this is a fictional example). In this table, there are three 2010 block groups: 061210001001, 061210001002, 061210003001, and two equivalence classes: one consisting of block group 061210001001 and the other consisting of block groups 061210001002 and 061210003001. A class (or equivalence class) is a set of block groups which share one or more blocks (they do not have to be the same block). Thus block group 061210001001 forms its own class because it has no shared blocks, while the other two block groups have one shared block (061210003003212). Classes are typified by the number of block groups in them, referred to as the degree of the class. For the two equivalence classes mentioned above, the first (consisting of block group 061210001001) has a degree of one while the second (consisting of block groups 061210001002 and 061210003001) has a degree of two.

It is useful to distinguish between those classes which have only one block group in them and those that have multiple block groups in them. This is because when there is only one block group in a class, a non-statistical comparison of the total

Table 32: 2010 block group conversion to 2020 blocks

2010 block group	2010 block	2020 block	Percent
061210001001	061210001001001	061210001002001	100.00
	061210001001002	061210001002003	100.00
	061210001001003	061210001002015	100.00
	061210001001004	061210001003002	100.00
061210001002	061210001002003	061210003002003	100.00
	061210001002008	061210003002005	100.00
	061210001002016	061210003002017	100.00
	061210001002024	061210003003212	50.00
061210003001	061210003001002	061210003003212	50.00
	061210003001004	061210003003213	100.00

citizen population from the CVAP data can be made to the total population from the P.L. 94-171 data.

If there are more citizens from the CVAP than total population from the P.L. 94-171 in the exact matching of a 2010 block group to 2020 geography (no split blocks between the 2010 block group and the 2020 geography), this indicates a mismatch of data between the two sources. To examine this, all 23,212 2010 block groups are classified by their equivalence class. The distribution of classes (by number of block groups) is given in table 33. As can be seen in this table, the vast majority (73.4%) of classes have only one block group in them. We can thus compare directly, without any statistical adjustments, the total population from the 2020 P.L. 94-171 and the citizen population from the 2015-2019 ACS.

For block groups with an equivalence class degree of one, then, we obtain (table 34) that nearly a quarter have more citizens from the ACS dataset than there are people in the block group, an impossibility. A further breakdown by the magnitudes (letting r = citizens/pop, pop = .000000001 when pop = 0), as displayed in table 35. There are also five block groups with a number of citizens greater than zero and zero population from the P.L. 94-171, and 16 with the zero citizens and non-zero population from the P.L. 94-171.

Recall that these are only 2010 block groups that are in a class by themselves (that is, there are no 2020 blocks split between that block group and another 2010 block group). Given the discrepancies between the ACS and P.L. 94-171, a straight-forward geographic allocation is made for those block groups which share one or more blocks with other block groups (the geographic allocation is based on proportion of the block falling within a block group). A very few adjustments are made to this geographic allocation method, but almost universally the geographic proportion is used for 2020 blocks split between 2010 block groups.

Table 33: 2010 block group conversion to 2020 block

block groups in class	number of classes	number of block groups
1	17045	17045
2	1161	2322
3	344	1032
4	169	676
5	80	400
6	64	384
7	31	217
8	17	136
9	16	144
10	4	40
11	5	55
12	7	84
13	3	39
14	6	84
15	4	60
16	2	32
17	1	17
18	4	72
19	3	57
20	4	80
21	1	21
22	2	44
23	3	69
29	1	29
32	1	32
41	1	41

Table 34: ACS Citizens versus P.L. 94-171 Population by unsplit block groups

classification	number
Citizens < Pop	12982
Citizens > Pop	4063

Table 35: ACS Citizens versus P.L. 94-171 Population by unsplit block groups, r = citizens/pop

r > 10	10 >= r > 5	5 >= r > 2	2 >= r > 1	1 >= r > .9	.9 >= r > .8	r >= .8
8	12	22	4021	3240	3633	6109

It should be noted that if an exact matching of block groups from 2010 to 2020 census geography (no geographic splitting of blocks) were desired, one could allocate the entire equivalence class at once and avoid the split blocks and geographical conversion. This would, however, make the geographic correspondence much worse for equivalence classes containing more than one block group, because the allocation would take place over a much larger geographic area. Given the many sources of uncertainty in these data sets already (as illustrated above), geographic splitting was chosen as the method of allocation for census blocks with multiple 2020 block groups listed in the conversion, rather than allocating an equivalence class in its entirety and avoiding split geography.

Appendices

Potential Geocoding Problems

For an illustration of potential geocoding problems, consider , reproduced in .

Figure 4: A 2010 Census Block Which is Not a City Block

The dark orange shaded portion is block 2008. Block 2008 is essentially Shoreline Drive (in Alameda). By the Census Bureau's criteria, the double lines that roughly form a square in the middle of this figure (which are Shoreline, Park, Broadway and Otis--see map below), should have been treated as a single block (with a water block in the middle--also see map below).

Now consider the address 2465 Shoreline Drive, Alameda, CA, which is an apartment complex that Google Maps puts at location A on the map below. The

Census Bureau's TIGER address/block equivalency files put that address into block 2008. The population, however, has actually been assigned to block 2012 in P.L. 94-171, so that geocoding has created a block with large registration (over 500 registered voters) and no population. On the other hand, block 2012 has a large population and no registered voters. This problem occurs because the address ranges released in the **TIGER** files do not agree with the Census Bureau's internal assignment files (which are not released). The primary source of block registration/population errors (zero population blocks with registration, blocks with population and no registration) come about from this type of mismatch.

Registration data

Statistical categories

Party

dem Party Democrat
rep Party Republican
aip Party American Independent
paf Party Peace and Freedom
misc Party Miscellaneous
lib Party Libertarian
nlp Party Natural Law
green Party Green
ref Party Reform (no longer recognized by state)
dcl Party Declined to State (that is, non stated at registration)

* *Parties which are not Democrat, Republican or Declined to State are Other*

Gender

male Male
female Female

Ethnicity

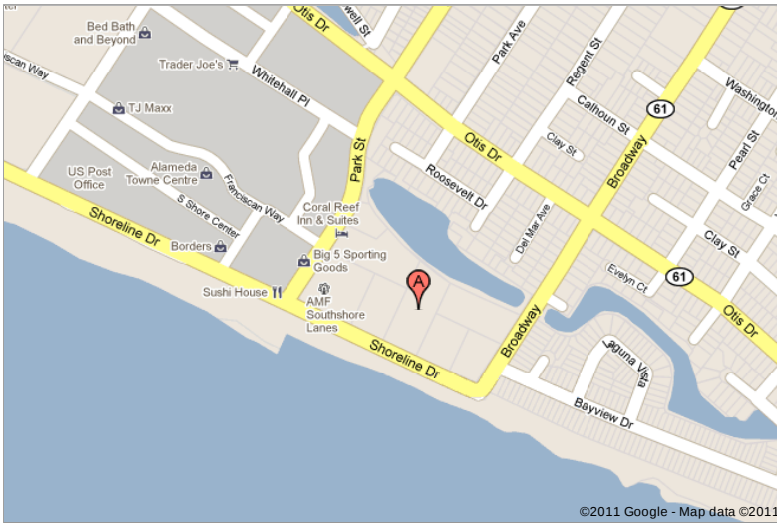
hispdem Latino Dems
hisprep Latino Reps
hispdcl Latino No Party
hispoth Latino Other Party

jewdem Jewish Dems
jewrep Jewish Reps
jewdcl Jewish No Party
jewoth Jewish Other Party

kordem Korean Dems
korrep Korean Reps
kordcl Korean No Party
koroth Korean Other Party

Google maps

To see all the details that are visible on the screen, use the "Print" link next to the map.



jpndem Japanese Dems
jpnrep Japanese Reps
jpndcl Japanese No Party
jpnoth Japanese Other Party

chidem Chinese Dems
chirep Chinese Reps
chidcl Chinese No Party
chioth Chinese Other Party

inddem Indian Dems
indrep Indian Reps
inddcl Indian No Party
indoth Indian Other Party

vietdem Vietnamese Dems
vietrep Vietnamese Reps vietdcl Vietnamese No Party
vietoth Vietnamese Other Party

fildem Filipino Dems
filrep Filipino Reps
fildcl Filipino No Party
filoth Filipino Other Party

Gender/Party/Age

male_dem_ageunk males dem birth date not listed on registered voter file
male_dem_age1824 males dem age between 18-24
male_dem_age2534 males dem age between 25-34
male_dem_age3544 males dem age between 35-44
male_dem_age4554 males dem age between 45-54
male_dem_age5564 males dem age between 55-64
male_dem_age65pl males dem age between 65 or older

female_dem_ageunk females dem age birth date not listed on registered voter file
female_dem_age1824 females dem age between 1824
female_dem_age2534 females dem age between 2534
female_dem_age3544 females dem age between 3544
female_dem_age4554 females dem age between 4554
female_dem_age5564 females dem age between 5564
female_dem_age65pl females dem age between 65 or older

male_rep_ageunk males rep age birth date not listed on registered voter file
male_rep_age1824 males rep age between 1824
male_rep_age2534 males rep age between 2534
male_rep_age3544 males rep age between 3544
male_rep_age4554 males rep age between 4554
male_rep_age5564 males rep age between 5564
male_rep_age65pl males rep age between 65 or older

female_rep_ageunk females rep age birth date not listed on registered voter file
female_rep_age1824 females rep age between 1824
female_rep_age2534 females rep age between 2534
female_rep_age3544 females rep age between 3544

female_rep_age4554 females rep age between 4554
female_rep_age5564 females rep age between 5564
female_rep_age65pl females rep age between 65 or older

male_dcl_ageunk males dcl age birth date not listed on registered voter file
male_dcl_age1824 males dcl age between 1824
male_dcl_age2534 males dcl age between 2534
male_dcl_age3544 males dcl age between 3544
male_dcl_age4554 males dcl age between 4554
male_dcl_age5564 males dcl age between 5564
male_dcl_age65pl males dcl age between 65 or older

female_dcl_ageunk females dcl age birth date not listed on registered voter file
female_dcl_age1824 females dcl age between 1824
female_dcl_age2534 females dcl age between 2534
female_dcl_age3544 females dcl age between 3544
female_dcl_age4554 females dcl age between 4554
female_dcl_age5564 females dcl age between 5564
female_dcl_age65pl females dcl age between 65 or older

male_oth_ageunk males oth age birth date not listed on registered voter file
male_oth_age1824 males oth age between 1824
male_oth_age2534 males oth age between 2534
male_oth_age3544 males oth age between 3544
male_oth_age4554 males oth age between 4554
male_oth_age5564 males oth age between 5564
male_oth_age65pl males oth age between 65 or older

female_oth_ageunk females oth age birth date not listed on registered voter file
female_oth_age1824 females oth age between 1824
female_oth_age2534 females oth age between 2534
female_oth_age3544 females oth age between 3544
female_oth_age4554 females oth age between 4554
female_oth_age5564 females oth age between 5564
female_oth_age65pl females oth age between 65 or older

Registration trends

dem_reg_cohort_1 Dems Registered after last general election
dem_reg_cohort_2 Dems Registered after 2nd to last general election
dem_reg_cohort_3 Dems Registered after 3rd to last general election
dem_reg_cohort_4 Dems Registered after 4th to last general election
dem_reg_cohort_5 Dems Registered after 5th to last general election
dem_reg_cohort_6 Dems Registered after 6th to last general election
dem_reg_cohort_7 Dems Registered after 7th to last general election
dem_reg_cohort_8 Dems Registered after 8th to last general election
dem_reg_cohort_9 Dems Registered anytime after 8th to last general election

rep_reg_cohort_1 Reps Registered after last general election

rep_reg_cohort_2 Reps Registered after 2nd to last general election
rep_reg_cohort_3 Reps Registered after 3rd to last general election
rep_reg_cohort_4 Reps Registered after 4th to last general election
rep_reg_cohort_5 Reps Registered after 5th to last general election
rep_reg_cohort_6 Reps Registered after 6th to last general election
rep_reg_cohort_7 Reps Registered after 7th to last general election
rep_reg_cohort_8 Reps Registered after 8th to last general election
rep_reg_cohort_9 Reps Registered anytime after 8th to last general election

dcl_reg_cohort_1 DCLs Registered after last general election
dcl_reg_cohort_2 DCLs Registered after 2nd to last general election
dcl_reg_cohort_3 DCLs Registered after 3rd to last general election
dcl_reg_cohort_4 DCLs Registered after 4th to last general election
dcl_reg_cohort_5 DCLs Registered after 5th to last general election
dcl_reg_cohort_6 DCLs Registered after 6th to last general election
dcl_reg_cohort_7 DCLs Registered after 7th to last general election
dcl_reg_cohort_8 DCLs Registered after 8th to last general election
dcl_reg_cohort_9 DCLs Registered anytime after 8th to last general election

oth_reg_cohort_1 Oths Registered after last general election
oth_reg_cohort_2 Oths Registered after 2nd to last general election
oth_reg_cohort_3 Oths Registered after 3rd to last general election
oth_reg_cohort_4 Oths Registered after 4th to last general election
oth_reg_cohort_5 Oths Registered after 5th to last general election
oth_reg_cohort_6 Oths Registered after 6th to last general election
oth_reg_cohort_7 Oths Registered after 7th to last general election
oth_reg_cohort_8 Oths Registered after 8th to last general election
oth_reg_cohort_9 Oths Registered anytime after 8th to last general election

List of Figures

List of Tables