# CREATING CALIFORNIA'S
# OFFICIAL REDISTRICTING DATABASE

Kenneth F. McCue, Ph.D.
Research Scientist, Department of Biology
California Institute of Technology

http://swdb.berkeley.edu/

August 2011

# I. TABLE OF CONTENTS

## II. EXECUTIVE SUMMARY

A database of population, ethnicity and political results is necessary for redistricting to ensure compliance with federal and state law. The Federal Voting Rights Act (VRA) is the most important of these, and that Act lays out criteria for determining whether a districting plan complies with the VRA. In order to determine whether a plan will comply with the VRA, it is necessary to look at voter registration and historic voting data.

Other legal criteria that plans must satisfy in California are laid out in Propositions 11 and 20. In addition, the Commission established under the procedures in Propositions 11 and 20 has wide latitude in making decisions on redistricting plans and may require data that allows descriptions of communities of interest, including economic and social characteristics revealed at least based in part by electoral, registration or census information.

The primary conceptual difficulty in creation of this database occurs because of the fact that much of the data that goes into this database is provided in aggregate form, and the data is from different and uncoordinated sources. Aggregate means that the data is not reported at the level of the individuals, but rather as an agglomeration over a (usually) contiguous geographic area. The different sources are the California county registrar of voters and county clerk offices and the United States Census Bureau. If all of 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 is not available at the individual level.

Breaking down the data into component sets, the first type is census data. Census data is collected at the individual level, but is 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 dataset for redistricting (PL94-171) 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 this data is collected at the individual level it is not reported that way, and just as the Census Bureau does not report individual data, there is no data reported at the level of the individual in the statewide database. The registration data collected on individuals is maintained by the California county registrar 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 on this file that is used by the statewide database for collecting statistics includes the voter partisan affiliation (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. This data is collected from each of the 58 county elections offices and is 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 registrar of voters and county clerk offices. This data is also collected individually (as are the census and registered voter information), but is only reported at the SOV precinct level. Data available are the votes on statewide constitutional offices, state legislative races, federal races, and statewide propositions.

These three types of data are thus reported by different geographic units. For redistricting purposes, the registration and electoral data needs 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 that 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 Spanish-surnamed voters). 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 by synonym tables; differences in street types can be handled by examining if 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 by 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 electoral results from the 2001 redistricting dataset or the citizen voting age population (CVAP) data. In both these cases, the data is in the 2000 census geography and must be converted into the 2010 census geography. The 2001 redistricting dataset is accurate at the census block level, and the conversion is relatively straightforward, using 2000 to 2010 census geography conversions created from the TIGER files. The CVAP data is at the level of the 2000 census block group level (there are typically 10 to 20 blocks in a block group), and furthermore, it is estimated from the ACS surveys, which are samples rather than complete enumeration. As a consequence, the level of error of the CVAP is higher than that of the 2000 elections results.

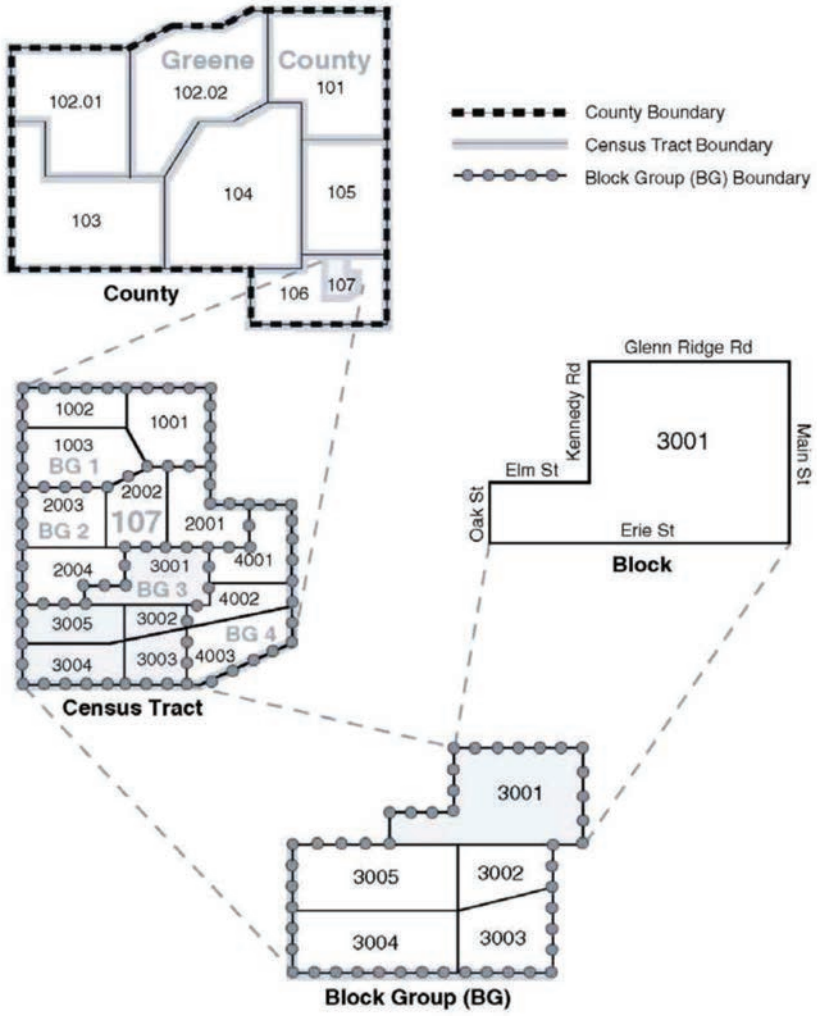
# III. INPUTS

### A. Census Data

Census data is 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.

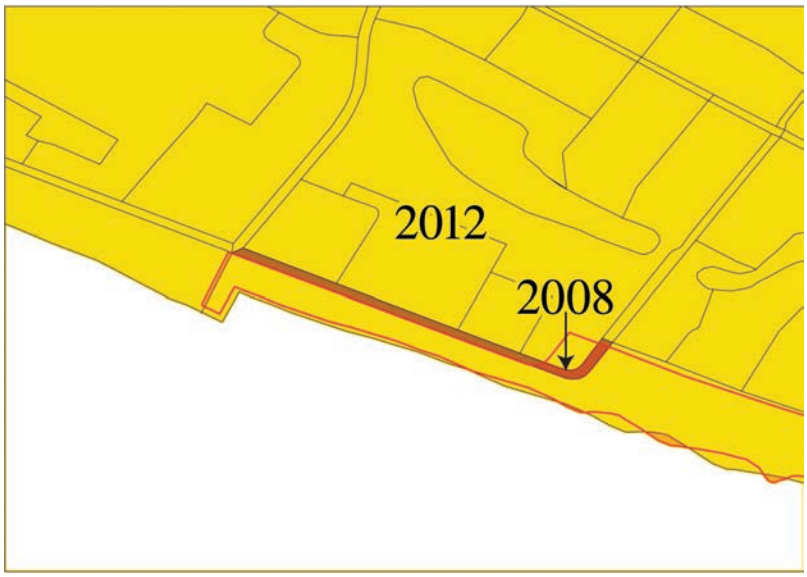
PL94-171 is the law under which population data is provided to the states for redistricting. The data in this law contains counts on population, both by age and by ethnic and racial membership. This data is reported by census block. A census block is typically a city block (though not always--this is discussed in the section on geocoding). This is illustrated in "***Figure 1: Geographic Relationships--Small Area Statistical Entities***", which is taken from the PL94-171 documentation (it is referred to as Figure 3 in that documentation). Counties are made up of census tracts which are made up of block groups which are made up of blocks (note that block groups are used in the CVAP [Citizen Voting-Age Population], which is discussed in the section on CVAP breakdown).

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



It should be noted that this is an idealized figure, as sometimes blocks are not city blocks. For example, in the following figure (***Figure 2)***, 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 VI).

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




The PL94-171 data is reported in five tables and is documented in 2010 Census Redistricting Data (Public Law 94-171) Summary File Technical Documentation/prepared by the U.S. Census Bureau, 2011. The PL94-171 file has five tables (one of which relates to housing). The population tables are: a count of all persons by race; a count of the population 18 years and over by race; a count of Hispanic or Latino and a count of not Hispanic or Latino by race for all persons; a count of Hispanic or Latino and a count of not Hispanic or Latino by race for the population 18 years and over. The race categories on all of these tables have multiple characterizations, because individuals can select membership in more than one race or ethnic group. The data in the PL94-171 is used in drawing districts.

The requirement of nearly equal populations in districts is expressed with the concept of the "ideal" number, which is the total number of people in a political entity divided by the number of districts. In California, for example, with a PL41-171 2011 population of 37,253,956, the ideal number for each Assembly district is 465,674.45 people. The race data is also used, in conjunction with political and registration data, to determine compliance with the Federal Voting Rights Act.

Additional technical documentation on the Public Law 94-171 Summary File can be found at ***http://www.census.gov/prod/cen2010/doc/pl94-171.pdf***. A description of the five tables is given in Appendix I.

In addition, a special tabulation of Citizen Voting Age Population Data (CVAP) was processed on the census block level so that it can be merged into the dataset.

**B. Registered Voter Data**

Registration voter data is information that the California county registrar of voters and county clerk offices maintain on individuals who are registered to vote. 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 affiliation is kept so that the registrar may ensure that only voters who have declared an affiliation with a party may vote in primary elections. The voters' date of registration is also kept, as voters may reregister at any time (to change party affiliation, for example). And, of course, the voter's name is included on this record. The information on the registered voter file is organized into certain categories in the statewide database, and counts are reported by various geographic units. One category is declared partisan affiliation. California has a minimum registration number requirement for listing as an officially recognized party, and there are a number of "minor" parties (so called because the number of adherents in those parties is relatively small) in addition to the Democratic and Republican party. In addition, there is a category called "Decline to State", which indicates that the individual decided not to state a party that the voter is affiliated with at time of registration. Other categories that are created from the information are those related to age (using the date of birth), passage of time since registration, and gender.

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 file 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 dictionary, the latter of which is 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: http://www.census.gov/population/www/documentation/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 straight last name to last name matching, various manipulations are made to surnames in order to fit them into the list.

Once the basic categories of partisan affiliation, age, ethnicity, time since registration and gender are created, variables which utilize two or more of these categories can be created. For example, the various ethnic/racial classifications are combined with partisan affiliation to create ethnicity by party. Age, sex and partisan affiliation are combined to create counts by, say, Dem males between 45 and 54 (inclusive). Partisan affiliations at the time of registration are used to create variables such as the number of decline to state voters who registered after the second to last general election.

Classes of these categories are then created so that each individual can be in one and only one category in each class. This avoids double counting, and it allows the aggregation of individuals to a unit of geography which accurately reflects the characteristics of the individual voters. For example, under this method, an individual can only be a member of one ethnicity, so the class of variables of ethnicity (which would be one of either Korean, Japanese, Chinese, Indian, Vietnamese, Filipino, Latino, or Jewish) can only be true for one of these variables. Thus, if in a census block with 40 registered voters, one has ten voters with a Latino surname, three with a Vietnamese surname, and eight with a Chinese surname, the aggregated values of these categories will reflect exactly that composition.

*The complete list of variables derived from the registered voter lists is given in Appendix II.*

**C. Election Data**

Election data in the statewide database include results from 12 statewide elections held in California over the last decade. These elections are the general elections in 2002, 2004, 2006, 2008 and 2010; the primary elections in 2004, 2006, 2008 and 2010; 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 and g10), primary elections similarly as p04, p06, p08 and p10, and the other elections have an s and the year (so s03 for the recall, s05 for the state-wide election called by Governor Schwarzenegger, and s08 for the presidential-only primary in February of 2008). In addition, three propositions from the 1990s, Propositions 187, 209 and 227, are included in the election results.

The races contained in the database (a full description is in Appendix III) consist of those pertaining to statewide constitutional offices, federal offices, state and federal legislative districts, and statewide propositions, initiatives and referenda. For general elections, all of these are present; for primaries, only competitive constitutional offices and federal offices are present, while state and federal legislative districts are not included for primaries. Only the top two candidates for any office are included. The merger of 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, 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 requiring that absentee voters in a particular physical precinct must also have their electoral results reported in a separate precinct, just as those who vote at the polls must have their election results reported in a separate precinct.

Before 2008, individuals who voted absentee often were 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. Individuals in different precincts will often get different ballots, as they may live in different legislative districts in a statewide election (or in different cities, if that city is holding an election simultaneously with the statewide election). After 2008, when the aforementioned law went into effect, such grouping together of absentee voting results was no longer allowed.

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 "Disaggregating 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 type 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.

*A complete list of election results is given in Appendix III.*

# IV. PROCESSING

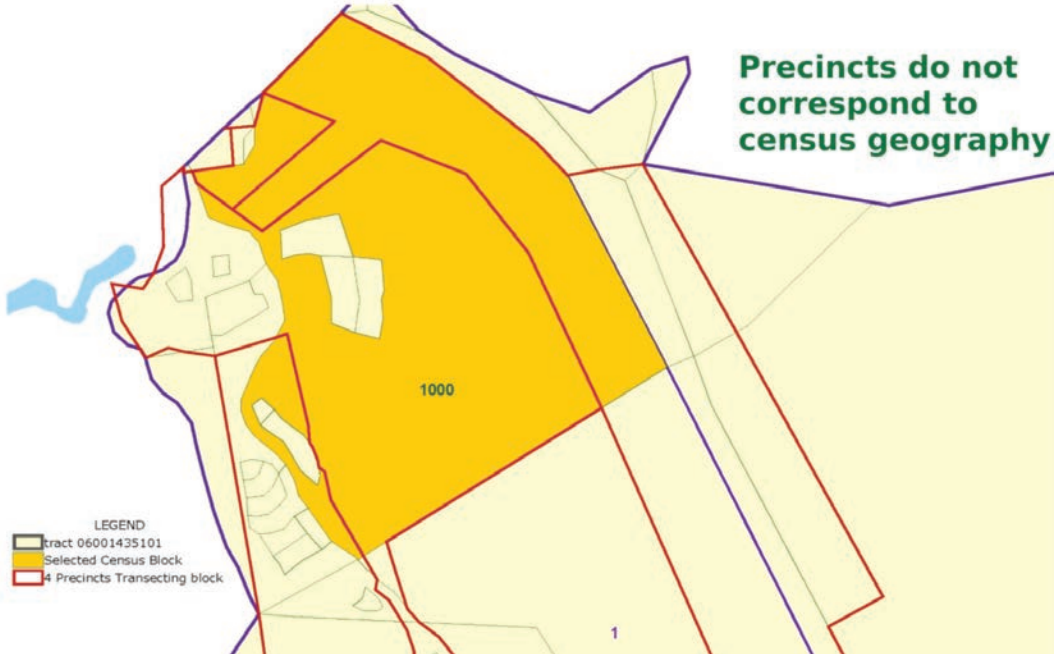
## A. Creating a Common Geography

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 is used for redistricting and election data at the voting precinct level is 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 SWDB. 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 that 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. ***Figure 3*** 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 described 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 is 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, and are described in detail below.

**B. 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 ``register'' (insure agreement) 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, along with the assignments made by geocoding, to allow the assignment of registered voters unassigned through geocoding to a census block. 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 $s$. 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, where the min is taken over all valid $u_{rb}$.

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.

***Table 1: Example of Registered Voter Assignments to Blocks***

| | 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 |

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 registration to population of blocks 200$i$. 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***.

**C. Converting Election Data to census geography**

The counties have registration (RG) and Voting or consolidated (SV) precincts as well as ballot groups for each election. Registration statistics are reported by RG precinct and election results are reported by SV precinct and/or ballot group (ABSPREC). The counties also report the grouping of RG precincts in SV precinct and Ballot group (although the ballot group reporting is inconsistent and incomplete).

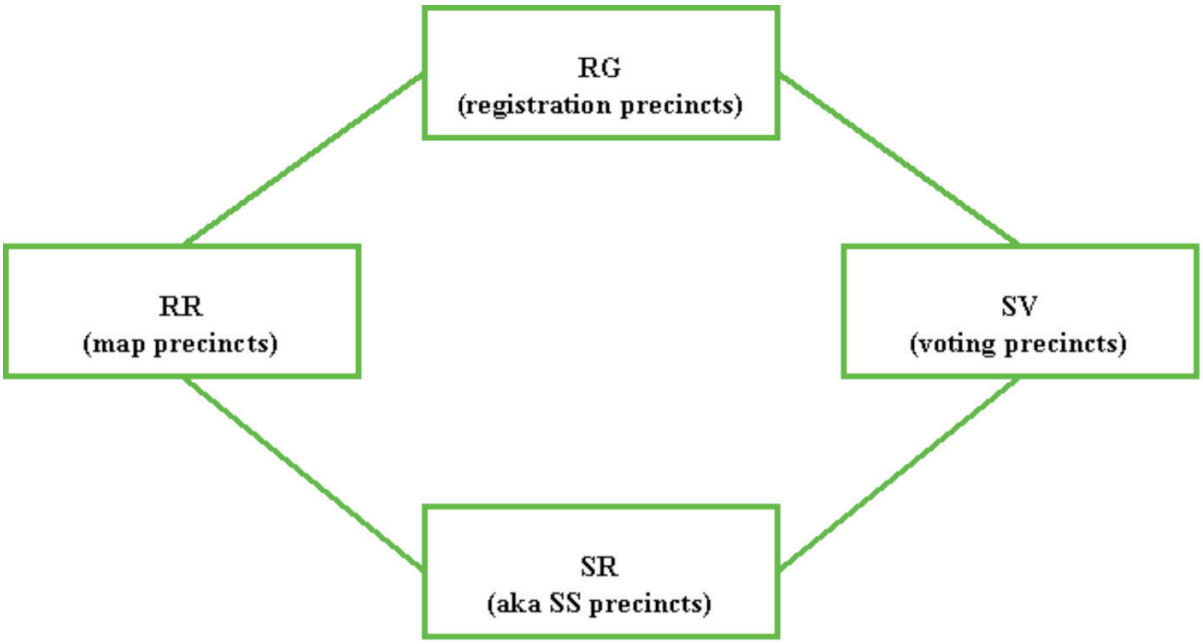
The counties also provide maps or GIS files showing registration precinct geography, but the geographic representations do not necessarily match the RG precincts for one or more of the following reasons:

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 SV precincts, we create a geographic consolidation known variously as SS or SR precincts to contain whole RR and SV 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.

***Figure 4: RG, RR, SV, SR precincts***




**D. 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 following is done to do this.

- 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 Left/Right/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, . . . , 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 estimate $p_{gi}$,:

$$\min \sum^{P} (v_i - \sum X_g p_{gi})^2,$$

where the min is taken over the $p_{gi}$, (in actual fact a more complicated optimization is used but it will produce results close to this equation).

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 $p_{gi}$, = $p_g(z_i, \mathcal{T})$, where $\mathcal{T}$ 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 race, each estimation should be done for each district by itself (or suitably modified, with the use of dummy variables), whereas statewide races can usually be estimated with all precincts.

- Adjust estimated probabilities to the precinct. The way of doing this is described in Appendix V. The method of estimation described there 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 $\hat{\rho}$ is estimate such that $v_i = \sum_g Xg_i \hat{\rho} g_i$. This $\hat{\rho} i$ (which is a $G$ by 1 vector) is estimated by

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

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

These $\hat{\rho} i$'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 $\hat{\rho}_A$ and those in precinct B will have an assigned voting propensity of $\hat{\rho}_B$.

As an example, consider a two-precinct district as described in Table 2 (in this table, quantities are suppressed for units that are not applicable). 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 many voters of each group 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 ($p_1$ for group $X_1$ and $p_2$ for group $X_2$) and then the adjusted probabilities ( $\hat{\rho}_1$ for group $X_1$ and $\hat{\rho}_2$ 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, suppose it is desired to find the votes for a candidate in census block 1002. This block is split between precinct A and precinct B. Then the candidate received 15.0 votes in precinct A, block 1002 from group 1, and 34.0 votes in precinct B, block 1002 from group 1, for a total of 49.0 votes from group 1. Similarly, the candidate received 65.0 plus 22.5 votes from group 2, from the part of block 1002 in precinct A and the part in block 1002 in precinct B, respectively. Thus there was a total of 146.5 votes for the candidate in census block 1002.

***Table 2: Example of Disaggregated Voting Probabilities***

| Precinct | v | $\hat{v}$ | Block | $X_1$ | $p_1$ | $\tilde{p}_1$ | $\hat{v}_1$ | $X_2$ | $p_2$ | $\tilde{p}_2$ | $\hat{v}_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 |

**E. Converting 2000 Census Geography to 2010 Census Geography**

While the techniques described so far are the most accurate method of merging political and registration data to census geography, there are times when it is not feasible to go through the processes described herein and other methods must be used. Typically that is when the data is not available in the same manner in which the data so far has been described. There are two types of data included in the statewide dataset which are of this type: the selected propositions from the 1990s and the CVAP (Citizen Voting Age Population) data. The methods for this type of merger are described below. Assuming that count data is in the form of 2000 census blocks, the conversion takes place as follows. First, a geographical conversion between 2000 and 2010 census geography is performed. This was done by overlaying the two shape files from the TIGER files and using a GIS (the particular one was QGIS) to calculate the intersections of the two geographies. Since this is a geographic conversion the actual area of the intersections is not used but rather a weighted sum of the 2010 populations for each 2000 block, the weight being made up of the proportion of population in each 2010 block which has an intersection with a 2000 block, that proportion being calculated from the population in the 2010 block divided by the total population of all 2010 blocks associated with the 2000 block. (If there is no population in any of the 2010 blocks, then equal proportions are assigned to all). As an example, let 2000 block 1001 be split between 2010 blocks 1001, 1002 and 1003, and assume that a data value of 30 is to be broken down from 2000 block 1001 to those three 2010 blocks.

***Table 3: 2000 to 2010 Data Conversions by Block***

| 2000 block | 2000 data | 2010 block | 2010 population | Proportion | 2010 data |
|---|---|---|---|---|---|
| 1001 | 30 | 1001 | 20 | .2 | 6 |
| | | 1002 | 50 | 5 | 15 |
| | | 1003 | 30 | .3 | 9 |
| | | | --- | --- | --- |
| | | Total | 100 | 1.0 | 30 |

So for the count data value of 30 associated with 2000 block 1001, .2(30) = 6 would be awarded to 2010 block 1001, .5(30)=15 would be awarded to 2010 block 1002, and .3(30) = 9 would be awarded to 2010 block 1003. While this can give fractional values, rounding constrained by no loss of data gives a 2000 value for each block. For the selected propositions, the process is done, since that data is in 2000 block form. For the CVAP data, which is given by block group, an additional step is necessary. Exactly as before, calculate weighted sums for the breakdown of the block group from the 2000 data, only this time the breakdown is from the 2000 census block group to the 2000 blocks associated with that block group, and the populations are 2000 block populations. For example, suppose 2000 block group 1 has blocks 1001, 1002, 1003, 1004 and 1005 with populations of 30, 20, 20, 0, and 10, respectively. Then to break down a value of 100, the following table can be constructed:

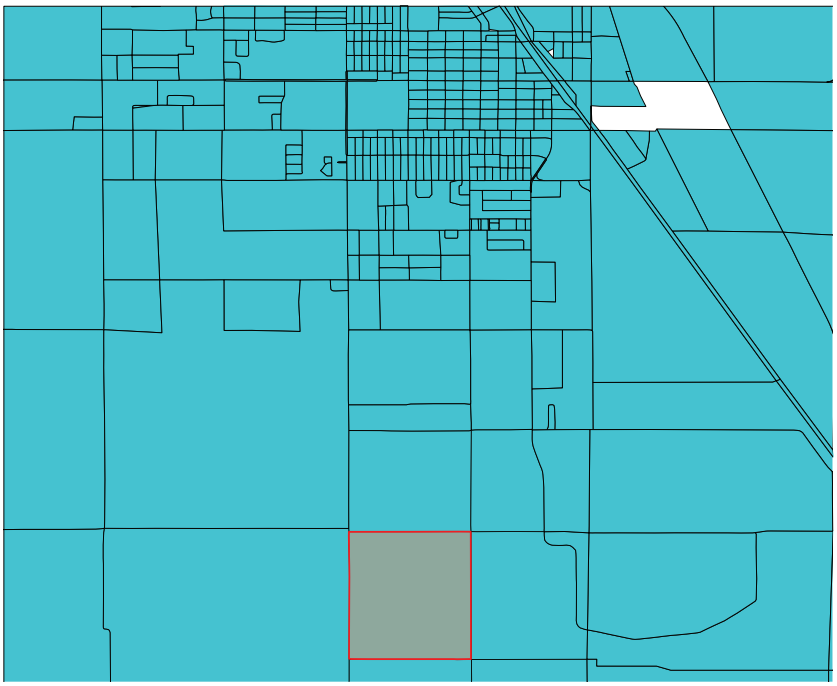
***Table 4: Converting 2000 block groups to 2000 blocks***

| 2000 block group | 2000 data | 2000 block | 2000 population | Proportion | 2000 data |
|---|---|---|---|---|---|
| 1001 | 100 | 1001 | 30 | .3 | 30 |
| | | 1002 | 20 | .2 | 20 |
| | | 1003 | 20 | .2 | 20 |
| | | 1004 | 0 | .0 | 0 |
| | | 1005 | 10 | .2 | 20 |
| | | | --- | --- | --- |
| | | Total | 100 | 1.0 | 100 |

Then the 2000 data in 2000 block 1001 can be used in the above algorithm. It should be noted that whatever error comes from this procedure, when block groups are used and broken down first, the error of the procedure is approximately squared, which is a great difference.

There are some problems with conversion from the 2000 blocks to 2010 census blocks, which are actually problems with the census data itself. A particularly egregious example is in King county, census tract 16.02 (this tract contains a housing development). In the 2000 census, 6,377 people were placed in block 1002 of this tract. In the 2010 census, 5,298 people were placed in block 1001 and 6,820 were placed in block 1006. The conversion from 2000 to 2010 puts block 1002 from the 2000 census into block 1003 of the 2010 census. The diagram below shows the geographic intersection between block 1002 from 2000 and blocks 1003 from 2010, and both blocks represent the same geographic area. Obviously, the population of block 1002 from 2000 did not move to block 1001 or 1006 in 2010 and depopulate block 1002. Rather, there was a misassignment in one of the two censuses. Thus, this problem is illustrated in the figure below.

***Figure 5: Coincident Geography for Census Block 1002 (2000 Census) and Census Block 1003 (2010 Census).***



Census block 1002 in the 2000 Census and Census block 1003 in the 2010 Census share the same geography, which is more darkly shaded in Figure 5.

***Table 5: Geographic conversions between 2000 census and 2010 census for tract 16.02, Kings County***

| 2000 blocks | 2010 blocks |
|---|---|
| 060310016021000 | 060310016021000 |
| 060310016021001 | 060310016021000 |
| 060310016021001 | 060310016021001 |
| 060310016021001 | 060310016021002 |
| 060310016021001 | 060310016021007 |
| 060310016021002 | 060310016021003 |
| 060310016021003 | 060310016021004 |
| 060310016021004 | 060310016021005 |
| 060310016021004 | 060310016021006 |
| 060310016021005 | 060310016021006 |

***Table 6: Population from the 2000 Census for tract 16.02, Kings County***

| 2000 blocks | 2000 Pop |
|---|---|
| 060310016021000 | 0 |
| 060310016021001 | 0 |
| 060310016021002 | 6377 |
| 060310016021003 | 0 |
| 060310016021004 | 0 |
| 060310016021005 | 0 |

***Table 7: Population from the 2010 Census for tract 16.02, Kings County***

| 2010 blocks | 2010 Pop |
|---|---|
| 060310016021000 | 0 |
| 060310016021001 | 5298 |
| 060310016021002 | 0 |
| 060310016021003 | 0 |
| 060310016021004 | 0 |
| 060310016021005 | 0 |
| 060310016021006 | 6820 |
| 060310016021007 | 0 |

It is clear that there can be large inaccuracies when using a geographic-based conversion from 2000 to 2010 geography. Since block groups are contained within tracts, and the Census Bureau attempts to maintain geographic continuity between census for census tracts, errors such as discussed for tract 16.02 will not tend to create problems when the unit being used is the whole census tract. So for use of whole census tracts for the 1990's proposition data, the errors discussed here should not be consequential. On the other hand, the CVAP data is not only geographical based in the 2000 census geography, the Census Bureau also specifically advises against using this data to provide a contemporaneous estimate of citizen voting-age population for use with the complete population counts as provided in the 2010 PL94-171 data. Thus while the CVAP data suffers from

the type of geographic conversion errors discussed in this section, there are also many other problems with using this data that have nothing to do with geographic conversion. This is discussed more fully in the next section.

**F. CVAP**

The CVAP (Citizen Voting Age Population) data is provided by census block group (refer to Figure 1 for how a block group relates to blocks) from the 2000 census and has two numbers for each of 13 groups. The two numbers are CVAP_EST and CVAP_MOE (for estimate and margin of error, respectively), and the tables (referred to as "lines"), are as follows:

***Table 8: Line Numbers and Titles for CVAP***

| LNNUMBER | LNTITLE |
|---|---|
| 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 |

As discussed before, CVAP is provided by the Census Bureau as a special tabulation from the 5-year ACS (American Community Survey) data. The accuracy of the ACS data is described in the Census Bureau publication, American Community Survey, Multiyear Accuracy of the Data (3-year 2007-2009 and 5-year 2005-2009). There are various formulas given in this publication for estimation of standard error but it is important to note that these formulas cannot be applied to the CVAP special tabulation as the estimate are given in terms of 2000 block groups, not 2010 block groups. The conversion from 2000 block groups to 2010 block groups results in another source of error into the data, one that cannot be quantified with the information given.

Another problem with the CVAP is that it is derived from a five-year data collection period. The Census Bureau's documentation cited above states: "Multiyear estimates cannot be used to say what is going on in any particular year in the period, only what the average value is over the full period." The PL94-171 data is based on a data collection period of early 2010, so it is a particular year. Using the CVAP data for comparison with that particular year is expressly discouraged by the Census Bureau ("[m]ultiyear estimates cannot be used to say what is going on in any particular year in the period").

Since the ACS data is based on 3-year and 5-year averages ending in 2009, it either underestimates, overestimates, or fails to count certain types of individuals. For example, if population is increasing in an area, the average will be below the population in 2010, whereas if population is decreasing, the average will be above the population in 2010. Also, those turning 18 in 2010 will not be counted as voting-age individuals, and those who are included in the ACS but die before 2010 will be counted. As older individuals are less likely to be citizens and younger individuals more likely (particularly in areas with larger number of non-citizens), this will create a tendency for an undercount of citizen voting-age population.

Another sort of undercount comes from the inaccuracies in the merger process of the CVAP data to the 2010 census geography. As described more fully in the 2000 to 2010 block conversion section, sometimes the conversions between 2000 geography and 2010 geography are not complete accurate. Typically this leads to data being placed in the wrong census block (but the correct census tract), but in some cases the geographic files provided by the Census Bureau (the TIGER files) do not even allow a geographic conversion. In those cases (and it typically happens in areas of rapid growth and relatively high non-citizen voting age populations), some population may be lost all together. The following two tables give the CVAP losses for the breakdown for 2000 block groups to 2000 blocks, and the CVAP losses for the breakdown from 2000 block groups to 2010 blocks.

***Table 9: Lost CVAP Population From Breakdown of 2000 Block Groups to 2000 Blocks By Line***

| Line | 2000 block group | 2000 blocks | diff |
|---|---|---|---|
| 1 | 21942932 | 21942920 | 12 |
| 2 | 16825797 | 16825774 | 23 |
| 3 | 130202 | 130202 | 0 |
| 4 | 2515156 | 2515153 | 3 |
| 5 | 1562755 | 1562754 | 1 |
| 6 | 77695 | 77695 | 0 |
| 7 | 12165152 | 12165148 | 4 |
| 8 | 120793 | 120793 | 0 |
| 9 | 111527 | 111527 | 0 |
| 10 | 49234 | 49234 | 0 |
| 11 | 24404 | 24404 | 0 |
| 12 | 67118 | 67118 | 0 |
| 13 | 5117162 | 5117161 | 1 |

Note the CVAP figures themselves are internally inconsistent--Line 2 (non-Latino) and Line 13 (Latino) should add to Line 1, but there is a difference of 27.

***Table 10: Lost CVAP Population from Breakdown from 2000 Block Groups to 2010 Blocks for Estimated CVAP Data, by Line***

| Line | 2000 block group | 2010 blocks | diff |
|---|---|---|---|
| 1 | 21942932 | 21936446 | 6486 |
| 2 | 16825797 | 16820227 | 5570 |
| 3 | 130202 | 130139 | 63 |
| 4 | 2515156 | 2514553 | 603 |
| 5 | 1562755 | 1562553 | 202 |
| 6 | 77695 | 77678 | 17 |
| 7 | 12165152 | 12160586 | 4566 |
| 8 | 120793 | 120756 | 37 |
| 9 | 111527 | 111494 | 33 |
| 10 | 49234 | 49228 | 6 |
| 11 | 24404 | 24401 | 3 |
| 12 | 67118 | 67102 | 16 |
| 13 | 5117162 | 5116238 | 924 |

## Kenneth F. McCue, Ph.D.

Dr. McCue is an expert in statistics and analysis of large datasets. He has worked extensively with population data and methods of estimation, and he supervised the construction of the California Statewide Database, maintained by the School of Law at the University of California, Berkeley. Since 1993, he has been responsible for merging relevant data at the electoral precinct and census block level and performing statistical analysis on the database.

# Appendices

## Appendix I: Census Data

**File 01—File Linking Fields** (comma delimited). These fields link File 01 with the geographic header and other files in the data set.

| Field name | Data dictionary Data type reference name | | Max size |
|---|---|---|---|
| File Identification<br>A/N | FILEID | | 6 |
| State/U.S. Abbreviation (USPS)<br>A | STUSAB | | 2 |
| Characteristic Iteration<br>A/N | CHARITER | | 3 |
| Characteristic Iteration File Sequence Number<br>A/N | CIFSN | | 2 |
| Logical Record Number<br>N | LOGRECNO | | 7 |
| **P1. RACE [71]** | | | |
| Universe: Total population | | | |
| Total: | P0010001 | 01 | 9 |
| Population of one race: | P0010002 | 01 | 9 |
| White alone | P0010003 | 01 | 9 |
| Black or African American alone | P0010004 | 01 | 9 |
| American Indian and Alaska Native alone | P0010005 | 01 | 9 |
| Asian alone | P0010006 | 01 | 9 |
| Native Hawaiian and Other Pacific Islander alone | P0010007 | 01 | 9 |
| Some Other Race alone | P0010008 | 01 | 9 |
| Two or More Races: | P0010009 | 01 | 9 |
| Population of two races: | P0010010 | 01 | 9 |
| White; Black or African American | P0010011 | 01 | 9 |
| White; American Indian and Alaska Native | P0010012 | 01 | 9 |
| White; Asian | P0010013 | 01 | 9 |
| White; Native Hawaiian and Other Pacific Islander | P0010014 | 01 | 9 |
| White; Some Other Race | P0010015 | 01 | 9 |
| Black or African American; American Indian and Alaska Native | P0010016 | 01 | 9 |
| Black or African American; Asian | P0010017 | 01 | 9 |
| Black or African American; Native Hawaiian and Other Pacific Islander | P0010018 | 01 | 9 |
| Black or African American; Some Other Race | P0010019 | 01 | 9 |
| American Indian and Alaska Native; Asian | P0010020 | 01 | 9 |
| American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0010021 | 01 | 9 |
| American Indian and Alaska Native; Some Other Race | P0010022 | 01 | 9 |
| Asian; Native Hawaiian and Other Pacific Islander | P0010023 | 01 | 9 |
| Asian; Some Other Race | P0010024 | 01 | 9 |
| Native Hawaiian and Other Pacific Islander; Some Other Race | P0010025 | 01 | 9 |
| Population of three races: | P0010026 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native | P0010027 | 01 | 9 |
| White; Black or African American; Asian | P0010028 | 01 | 9 |
| White; Black or African American; Native Hawaiian and Other Pacific Islander | P0010029 | 01 | 9 |
| White; Black or African American; Some Other Race | P0010030 | 01 | 9 |
| White; American Indian and Alaska Native; Asian | P0010031 | 01 | 9 |
| White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0010032 | 01 | 9 |
| White; American Indian and Alaska Native; Some Other Race | P0010033 | 01 | 9 |
| White; Asian; Native Hawaiian and Other Pacific Islander | P0010034 | 01 | 9 |
| White; Asian; Some Other Race | P0010035 | 01 | 9 |
| White; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010036 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian | P0010037 | 01 | 9 |

| | | | |
|---|---|---|---|
| Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0010038 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Some Other Race | P0010039 | 01 | 9 |
| Black or African American; Asian; Native Hawaiian and Other Pacific Islander | P0010040 | 01 | 9 |
| Black or African American; Asian; Some Other Race | P0010041 | 01 | 9 |
| Black or African American; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010042 | 01 | 9 |
| American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0010043 | 01 | 9 |
| American Indian and Alaska Native; Asian; Some Other Race | P0010044 | 01 | 9 |
| American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010045 | 01 | 9 |
| Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010046 | 01 | 9 |
| Population of four races: | P0010047 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian | P0010048 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0010049 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Some Other Race | P0010050 | 01 | 9 |
| White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander | P0010051 | 01 | 9 |
| White; Black or African American; Asian; Some Other Race | P0010052 | 01 | 9 |
| White; Black or African American; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010053 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0010054 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Some Other Race | P0010055 | 01 | 9 |
| White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010056 | 01 | 9 |
| White; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010057 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0010058 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Some Other Race | P0010059 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010060 | 01 | 9 |
| Black or African American; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010061 | 01 | 9 |
| American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010062 | 01 | 9 |
| Population of five races: | P0010063 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0010064 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Some Other Race | P0010065 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010066 | 01 | 9 |
| White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010067 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010068 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010069 | 01 | 9 |
| Population of six races: | P0010070 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0010071 | 01 | 9 |

**P2. HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE [73]**

*Universe: Total population*

| | | | |
|---|---|---|---|
| Total: | P0020001 | 01 | 9 |
| Hispanic or Latino | P0020002 | 01 | 9 |
| Not Hispanic or Latino: | P0020003 | 01 | 9 |
| Population of one race: | P0020004 | 01 | 9 |
| White alone | P0020005 | 01 | 9 |
| Black or African American alone | P0020006 | 01 | 9 |
| American Indian and Alaska Native alone | P0020007 | 01 | 9 |
| Asian alone | P0020008 | 01 | 9 |
| Native Hawaiian and Other Pacific Islander alone | P0020009 | 01 | 9 |
| Some Other Race alone | P0020010 | 01 | 9 |
| Two or More Races: | P0020011 | 01 | 9 |
| Population of two races: | P0020012 | 01 | 9 |
| White; Black or African American | P0020013 | 01 | 9 |
| White; American Indian and Alaska Native | P0020014 | 01 | 9 |
| White; Asian | P0020015 | 01 | 9 |
| White; Native Hawaiian and Other Pacific Islander | P0020016 | 01 | 9 |
| White; Some Other Race | P0020017 | 01 | 9 |
| Black or African American; American Indian and Alaska Native | P0020018 | 01 | 9 |
| Black or African American; Asian | P0020019 | 01 | 9 |
| Black or African American; Native Hawaiian and Other Pacific Islander | P0020020 | 01 | 9 |
| Black or African American; Some Other Race | P0020021 | 01 | 9 |
| American Indian and Alaska Native; Asian | P0020022 | 01 | 9 |
| American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0020023 | 01 | 9 |
| American Indian and Alaska Native; Some Other Race | P0020024 | 01 | 9 |
| Asian; Native Hawaiian and Other Pacific Islander | P0020025 | 01 | 9 |
| Asian; Some Other Race | P0020026 | 01 | 9 |
| Native Hawaiian and Other Pacific Islander; Some Other Race | P0020027 | 01 | 9 |
| Population of three races: | P0020028 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native | P0020029 | 01 | 9 |
| White; Black or African American; Asian | P0020030 | 01 | 9 |
| White; Black or African American; Native Hawaiian and Other Pacific Islander | P0020031 | 01 | 9 |
| White; Black or African American; Some Other Race | P0020032 | 01 | 9 |
| White; American Indian and Alaska Native; Asian | P0020033 | 01 | 9 |
| White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0020034 | 01 | 9 |
| White; American Indian and Alaska Native; Some Other Race | P0020035 | 01 | 9 |
| White; Asian; Native Hawaiian and Other Pacific Islander | P0020036 | 01 | 9 |
| White; Asian; Some Other Race | P0020037 | 01 | 9 |
| White; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020038 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian | P0020039 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0020040 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Some Other Race | P0020041 | 01 | 9 |
| Black or African American; Asian; Native Hawaiian and Other Pacific Islander | P0020042 | 01 | 9 |
| Black or African American; Asian; Some Other Race | P0020043 | 01 | 9 |
| Black or African American; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020044 | 01 | 9 |
| American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0020045 | 01 | 9 |
| American Indian and Alaska Native; Asian; Some Other Race | P0020046 | 01 | 9 |
| American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020047 | 01 | 9 |
| Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020048 | 01 | 9 |

| | | | |
|---|---|---|---|
| Population of four races: | P0020049 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian | P0020050 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0020051 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Some Other Race | P0020052 | 01 | 9 |
| White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander | P0020053 | 01 | 9 |
| White; Black or African American; Asian; Some Other Race | P0020054 | 01 | 9 |
| White; Black or African American; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020055 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0020056 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Some Other Race | P0020057 | 01 | 9 |
| White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020058 | 01 | 9 |
| White; Asian; Native Hawaiian and Other Pacific Islander, Some Other Race | P0020059 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0020060 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Some Other Race | P0020061 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020062 | 01 | 9 |
| Black or African American; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020063 | 01 | 9 |
| American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020064 | 01 | 9 |
| Population of five races: | P0020065 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0020066 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Some Other Race | P0020067 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020068 | 01 | 9 |
| White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020069 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020070 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020071 | 01 | 9 |
| Population of six races: | P0020072 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0020073 | 01 | 9 |

**File 02—File Linking Fields** (comma delimited). These fields link File 02 with the geographic header and other files in the data set.

| Field name | Data dictionary Data type reference name | | Max size |
|---|---|---|---|
| File Identification<br>A/N | FILEID | | 6 |
| State/U.S. Abbreviation (USPS)<br>A | STUSAB | | 2 |
| Characteristic Iteration<br>A/N | CHARITER | | 3 |
| Characteristic Iteration File Sequence Number<br>A/N | CIFSN | | 2 |
| Logical Record Number<br>N | LOGRECNO | | 7 |
| **P3. RACE FOR THE POPULATION 18 YEARS AND OVER [71]** | | | |
| *Universe: Total population 18 years and over* | | | |
| Total: | P0030001 | 01 | 9 |
| Population of one race: | P0030002 | 01 | 9 |
| White alone | P0030003 | 01 | 9 |
| Black or African American alone | P0030004 | 01 | 9 |
| American Indian and Alaska Native alone | P0030005 | 01 | 9 |
| Asian alone | P0030006 | 01 | 9 |
| Native Hawaiian and Other Pacific Islander alone | P0030007 | 01 | 9 |
| Some Other Race alone | P0030008 | 01 | 9 |
| Two or More Races: | P0030009 | 01 | 9 |
| Population of two races: | P0030010 | 01 | 9 |
| White; Black or African American | P0030011 | 01 | 9 |
| White; American Indian and Alaska Native | P0030012 | 01 | 9 |
| White; Asian | P0030013 | 01 | 9 |
| White; Native Hawaiian and Other Pacific Islander | P0030014 | 01 | 9 |
| White; Some Other Race | P0030015 | 01 | 9 |
| Black or African American; American Indian and Alaska Native | P0030016 | 01 | 9 |
| Black or African American; Asian | P0030017 | 01 | 9 |
| Black or African American; Native Hawaiian and Other Pacific Islander | P0030018 | 01 | 9 |
| Black or African American; Some Other Race | P0030019 | 01 | 9 |
| American Indian and Alaska Native; Asian | P0030020 | 01 | 9 |
| American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0030021 | 01 | 9 |
| American Indian and Alaska Native; Some Other Race | P0030022 | 01 | 9 |
| Asian; Native Hawaiian and Other Pacific Islander | P0030023 | 01 | 9 |
| Asian; Some Other Race | P0030024 | 01 | 9 |
| Native Hawaiian and Other Pacific Islander; Some Other Race | P0030025 | 01 | 9 |
| Population of three races: | P0030026 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native | P0030027 | 01 | 9 |
| White; Black or African American; Asian | P0030028 | 01 | 9 |
| White; Black or African American; Native Hawaiian and Other Pacific Islander | P0030029 | 01 | 9 |
| White; Black or African American; Some Other Race | P0030030 | 01 | 9 |
| White; American Indian and Alaska Native; Asian | P0030031 | 01 | 9 |
| White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0030032 | 01 | 9 |
| White; American Indian and Alaska Native; Some Other Race | P0030033 | 01 | 9 |
| White; Asian; Native Hawaiian and Other Pacific Islander | P0030034 | 01 | 9 |
| White; Asian; Some Other Race | P0030035 | 01 | 9 |
| White; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030036 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian | P0030037 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0030038 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Some Other Race | P0030039 | 01 | 9 |

| | | | |
|---|---|---|---|
| Black or African American; Asian; Native Hawaiian and Other Pacific Islander | P0030040 | 01 | 9 |
| Black or African American; Asian; Some Other Race | P0030041 | 01 | 9 |
| Black or African American; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030042 | 01 | 9 |
| American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0030043 | 01 | 9 |
| American Indian and Alaska Native; Asian; Some Other Race | P0030044 | 01 | 9 |
| American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030045 | 01 | 9 |
| Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030046 | 01 | 9 |
| Population of four races: | P0030047 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian | P0030048 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0030049 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Some Other Race | P0030050 | 01 | 9 |
| White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander | P0030051 | 01 | 9 |
| White; Black or African American; Asian; Some Other Race | P0030052 | 01 | 9 |
| White; Black or African American; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030053 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0030054 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Some Other Race | P0030055 | 01 | 9 |
| White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030056 | 01 | 9 |
| White; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030057 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0030058 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Some Other Race | P0030059 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030060 | 01 | 9 |
| Black or African American; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030061 | 01 | 9 |
| American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030062 | 01 | 9 |
| Population of five races: | P0030063 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0030064 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Some Other Race | P0030065 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030066 | 01 | 9 |
| White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030067 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030068 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030069 | 01 | 9 |
| Population of six races: | P0030070 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0030071 | 01 | 9 |

**P4. HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE FOR THE POPULATION 18 YEARS AND OVER [73]**

*Universe: Total population 18 years and over*

| | | | |
|---|---|---|---|
| Total: | P0040001 | 01 | 9 |
| Hispanic or Latino | P0040002 | 01 | 9 |
| Not Hispanic or Latino: | P0040003 | 01 | 9 |
| Population of one race: | P0040004 | 01 | 9 |
| White alone | P0040005 | 01 | 9 |
| Black or African American alone | P0040006 | 01 | 9 |
| American Indian and Alaska Native alone | P0040007 | 01 | 9 |
| Asian alone | P0040008 | 01 | 9 |
| Native Hawaiian and Other Pacific Islander alone | P0040009 | 01 | 9 |
| Some Other Race alone | P0040010 | 01 | 9 |
| Two or More Races: | P0040011 | 01 | 9 |
| Population of two races: | P0040012 | 01 | 9 |
| White; Black or African American | P0040013 | 01 | 9 |
| White; American Indian and Alaska Native | P0040014 | 01 | 9 |
| White; Asian | P0040015 | 01 | 9 |
| White; Native Hawaiian and Other Pacific Islander | P0040016 | 01 | 9 |
| White; Some Other Race | P0040017 | 01 | 9 |
| Black or African American; American Indian and Alaska Native | P0040018 | 01 | 9 |
| Black or African American; Asian | P0040019 | 01 | 9 |
| Black or African American; Native Hawaiian and Other Pacific Islander | P0040020 | 01 | 9 |
| Black or African American; Some Other Race | P0040021 | 01 | 9 |
| American Indian and Alaska Native; Asian | P0040022 | 01 | 9 |
| American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0040023 | 01 | 9 |
| American Indian and Alaska Native; Some Other Race | P0040024 | 01 | 9 |
| Asian; Native Hawaiian and Other Pacific Islander | P0040025 | 01 | 9 |
| Asian; Some Other Race | P0040026 | 01 | 9 |
| Native Hawaiian and Other Pacific Islander; Some Other Race | P0040027 | 01 | 9 |
| Population of three races: | P0040028 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native | P0040029 | 01 | 9 |
| White; Black or African American; Asian | P0040030 | 01 | 9 |
| White; Black or African American; Native Hawaiian and Other Pacific Islander | P0040031 | 01 | 9 |
| White; Black or African American; Some Other Race | P0040032 | 01 | 9 |
| White; American Indian and Alaska Native; Asian | P0040033 | 01 | 9 |
| White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0040034 | 01 | 9 |
| White; American Indian and Alaska Native; Some Other Race | P0040035 | 01 | 9 |
| White; Asian; Native Hawaiian and Other Pacific Islander | P0040036 | 01 | 9 |
| White; Asian; Some Other Race | P0040037 | 01 | 9 |
| White; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040038 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian | P0040039 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0040040 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Some Other Race | P0040041 | 01 | 9 |
| Black or African American; Asian; Native Hawaiian and Other Pacific Islander | P0040042 | 01 | 9 |
| Black or African American; Asian; Some Other Race | P0040043 | 01 | 9 |
| Black or African American; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040044 | 01 | 9 |
| American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0040045 | 01 | 9 |
| American Indian and Alaska Native; Asian; Some Other Race | P0040046 | 01 | 9 |
| American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040047 | 01 | 9 |
| Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040048 | 01 | 9 |

| | | | |
|---|---|---|---|
| Population of four races: | P0040049 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian | P0040050 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander | P0040051 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Some Other Race | P0040052 | 01 | 9 |
| White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander | P0040053 | 01 | 9 |
| White; Black or African American; Asian; Some Other Race | P0040054 | 01 | 9 |
| White; Black or African American; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040055 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0040056 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Some Other Race | P0040057 | 01 | 9 |
| White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040058 | 01 | 9 |
| White; Asian; Native Hawaiian and Other Pacific Islander, Some Other Race | P0040059 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0040060 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Some Other Race | P0040061 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040062 | 01 | 9 |
| Black or African American; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040063 | 01 | 9 |
| American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040064 | 01 | 9 |
| Population of five races: | P0040065 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander | P0040066 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Some Other Race | P0040067 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040068 | 01 | 9 |
| White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040069 | 01 | 9 |
| White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040070 | 01 | 9 |
| Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040071 | 01 | 9 |
| Population of six races: | P0040072 | 01 | 9 |
| White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander; Some Other Race | P0040073 | 01 | 9 |

**H1. OCCUPANCY STATUS [3]**

*Universe: Housing units*

| | | | |
|---|---|---|---|
| Total: | H0010001 | 02 | 9 |
| Occupied | H0010002 | 02 | 9 |
| Vacant | H0010003 | 02 | 9 |

## Appendix II: 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 |
| dcl | Party Declined to State (that is, non stated at registration) |

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

**Sex**

| | |
|---|---|
| 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 |
| 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 |

**Sex/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 |

## Appendix III: Election data

| Election | Office | Dem | Rep |
|---|---|---|---|
| g02 | GOV | Davis | Simon |
| g02 | LTG | Bustamante | McPherson |
| g02 | SOS | Shelley | Olberg |
| g02 | CON | Westley | McClintock |
| g02 | TRS | Angelides | Conlon |
| g02 | ATG | Lockyer | Ackerman |
| g02 | INS | Garamendi | Mendoza |
| g02 | CNG | district candidate* | district candidate* |
| g02 | SEN | district candidate* | district candidate* |
| g02 | ASS | district candidate* | district candidate* |

| Election | Proposition | Title |
|---|---|---|
| g02 | PR_46 (Y and N) | Housing and Emergency Shelter Trust Fund Act of 2002 |
| g02 | PR_47 (Y and N) | Kindergarten-University Public Education Facilities Bond Act of 2002 |
| g02 | PR_48 (Y and N) | Court Consolidation |
| g02 | PR_49 (Y and N) | Before and After School Programs. State Grants. |
| g02 | PR_50 (Y and N) | Water Quality, Supply and Safe drinking Water Projects. Coastal Wetlands Purchase and Protection |
| g02 | PR_51 (Y and N) | Transportation. Distribution of Existing Motor Vehicle Sales and Use Tax |
| g02 | PR_52 (Y and N) | Election Day Voter Registration. Voter Fraud Penalties. |

| Election | Office | Dem | Rep |
|---|---|---|---|
| s03 | GOV | Bustamante | Schwarzenegger |

| Election | Proposition | Title |
|---|---|---|
| s03 | PR_53 (Y and N) | Funds Dedicated for State and Local Infrastucture |
| s03 | PR_54 (Y and N) | Classification by Race, Ethnicity, Color, or National Origin |

| Election | Proposition | Title |
|---|---|---|
| p04 | PR_55(Y and N) | Education Bond |
| p04 | PR_56 (Y and N) | State Budget |
| p04 | PR_57 (Y and N) | Economic Recovery Bond |
| p04 | PR_58 (Y and N) | Balanced Budget |

| Election | Office | Dem 1 | Dem 2 |
|---|---|---|---|
| p04 | USS | Marin | Jones |

| Election | Office | Dem | Rep |
|---|---|---|---|
| g04 | PRS | Kerry | Bush |
| g04 | USS | Boxer | Jones |

| Election | Proposition | Title |
|---|---|---|
| g04 | PR_1A (Y and N) | Protection of Local Government Revenues |
| g04 | PR_59 (Y and N) | Public Records. Open Meetings |
| g04 | PR_60 (Y and N) | Election Rights of Political Parties |
| g04 | PR_60A (Y and N) | Surplus Property |
| g04 | PR_61 (Y and N) | Children's Hospital Projects. Grant Program |
| g04 | PR_62 (Y and N) | Elections. Primaries |
| g04 | PR_63 (Y and N) | Mental Health Services Expansion. Funding. Tax on Personal Incomes above $1 million. |

* The specific top-two Democratic and Republican candidates in the State Assembly, Congressional and Senate district races varies by district.

| | | |
|---|---|---|
| g04 | PR_64 (Y and N) | Limits on Private Enforcement of Unfair Business Competition Laws. |
| g04 | PR_65 (Y and N) | Local Government Funds, Revenues. State Mandates. |
| g04 | PR_66 (Y and N) | Limitations on "Three Strikes" Law. Sex Crimes. Punishment. |
| g04 | PR_67 (Y and N) | Emergency Medical Services. Funding. Telephone Surcharge. |
| g04 | PR_68 (Y and N) | Non-Tribal Commercial Gambling Expansion. Tribal Gaming Compact Amendments |
| g04 | PR_69 (Y and N) | DNA Samples. Collection. Database. Funding |
| g04 | PR_70 (Y and N) | Tribal Gaming Compacts. Exclusive Gaming Rights. Contributions to State |
| g04 | PR_71 (Y and N) | Stem Cell Research |
| g04 | PR_72 (Y and N) | Health Care Coverage Requirements |

| Election | Proposition | Title |
|---|---|---|
| s05 | PR_73 (Y and N) | Waiting Period and Parental Notification Before Termination of Minor's Pregnancy |
| s05 | PR_74 (Y and N) | Public School Teachers. Waiting Period for Permanent Status. Dismissal |
| s05 | PR_75 (Y and N) | Public Employee Union Dues, Restrictions on Political Contributions |
| s05 | PR_76 (Y and N) | State Spending and School Funding Limits. Initiative Constitutional Amendment |
| s05 | PR_77 (Y and N) | Redistricting Initiative Constitutional Amendment |
| s05 | PR_78 (Y and N) | Discounts on Prescription Drugs. Initiative Statute |
| s05 | PR_79 (Y and N) | Prescription Drug Discounts. State-negotiated Rebates. |
| s05 | PR_80 (Y and N) | Electric Service Providers. Regulation. Initiative Statute |

| Election | Office | Dem 1 | Dem 2 |
|---|---|---|---|
| p06 | GOV | Angelides | Westly |
| p06 | LTG | Garamendi | Figueroa |
| p06 | SOS | Bowen | Ortiz |

| Election | Office | Rep 1 | Rep 2 |
|---|---|---|---|
| p06 | LTG | McClintock | Farmer |

| Election | Office | Dem 1 | Dem 2 |
|---|---|---|---|
| p06 | CON | Dunn | Chiang |

| Election | Office | Rep 1 | Rep 2 |
|---|---|---|---|
| p06 | CON | Harris | Strickland |
| p06 | TRS | Parrish | Richman |

| Election | Office | Dem 1 | Dem 2 |
|---|---|---|---|
| p06 | ATG | Delgadillo | Brown |
| p06 | INS | Bustamante | Kraft |

| Election | Office | Rep 1 |
|---|---|---|
| p06 | USS | Mountjoy |

| Election | Proposition | Title |
|---|---|---|
| p06 | PR_81 (Y and N) | Reading Improvement, Library Renovation Bond Act |
| p06 | PR_82 (Y and N) | Preschool Education. Tax Increase Upper Income. |

| Election | Office | Dem | Rep |
|---|---|---|---|
| g06 | GOV | Angelides | Schwarzenegger |
| g06 | LTG | Garamendi | McClintock |
| g06 | SOS | Bowen | McPherson |
| g06 | CON | Chiang | Strickland |
| g06 | INS | Bustamante | Poizner |
| g06 | TRS | Lockyer | Parrish |
| g06 | ATG | Brown | Poochigian |
| g06 | USS | Feinstein | Mountjoy |
| g06 | CNG | district candidate* | district candidate* |
| g06 | SEN | district candidate* | district candidate* |
| g06 | ASS | district candidate* | district candidate* |

| Election | Proposition | Title |
|---|---|---|
| g06 | PR_1A (Y and N) | Transportation Funding Protection |
| g06 | PR_1B (Y and N) | Highway Safety/Air Quality/Port Security Bond 2006 |
| g06 | PR_1C (Y and N) | Housing/Emergency Shelter Trust Fund 2006 |
| g06 | PR_1D (Y and N) | Public Education Facilities bond 2006 |
| g06 | PR_1E (Y and N) | Disaster Preparedness/Flood Prevention Bond 2006 |
| g06 | PR_83 (Y and N) | Sex Offenders/Residence Restrictions Monitoring |

* The specific top-two Democratic and Republican candidates in the State Assembly, Congressional and Senate district races varies by district.

| | | |
|---|---|---|
| g06 | PR_84 (Y and N) | Water Qual/Flood/Resource Protection/Park Bonds |
| g06 | PR_85 (Y and N) | Waiting Period/Parental Notification |
| g06 | PR_86 (Y and N) | Cigarette Tax Initiative |
| g06 | PR_87 (Y and N) | Alternative Energy/Research/Oil Producer Tax |
| g06 | PR_88 (Y and N) | Education Funding/Real Property Parcel Tax |
| g06 | PR_89 (Y and N) | Political Campaigns/Public Financing/Corp Tax |
| g06 | PR_90 (Y and N) | Govt Acquisition/Regulation of Private Property |

| Election | Proposition | Title |
|---|---|---|
| s08 | PR_91 (Y and N) | Transportation Funds |
| s08 | PR_92 (Y and N) | Community Colleges. Funding. Governance. Fees. |
| s08 | PR_93 (Y and N) | Limits on Legislators' Terms of Office |
| s08 | PR_94 (Y and N) | Ref on Amd to Indian Gaming compact. Pechanga |
| s08 | PR_95 (Y and N) | Ref on Amd to Indian Gaming Compact. Morongo |
| s08 | PR_96 (Y and N) | Ref on Amd to Indian Gaming Compact. Sycuan |
| s08 | PR_97 (Y and N) | Ref on Amd to Indian Gaming Compact. Agua Caliente |

| Election | Office | Dem 1 | Dem 2 |
|---|---|---|---|
| s08 | PRESDEM | Obama | Clinton |

| Election | Office | Rep 1 | Rep 2 |
|---|---|---|---|
| s08 | PRESREP | McCain | Romney |

| Election | Proposition | Title |
|---|---|---|
| p08 | PR_98 (Y and N) | Eminent Domain. Limits on Government Authority. Initiative Constitutional Amendment |
| p08 | PR_99 (Y and N) | Eminent Domain. Limits on Government Acquisition of Owner-Occupied Residence. Initiative Constitutional Amendment |

| Election | Office | Dem | Rep |
|---|---|---|---|
| g08 | PRS | Obama | McCain |
| g08 | SEN | district candidate* | district candidate* |
| g08 | ASS | district candidate* | district candidate* |

| Election | Proposition | Title |
|---|---|---|
| g08 | PR_10 (Y and N) | Alternative Fuel Vehicles and renewable Energy Bonds |
| g08 | PR_11 (Y and N) | Redistricting |
| g08 | PR_12 (Y and N) | Veterans' Bonds |
| g08 | PR_1A (Y and N) | Safe, Reliable High-Speed Train Bond Act |
| g08 | PR_5 (Y and N) | Nonviolent Drug Offense, Sentencing, Parole, Rehab |
| g08 | PR_6 (Y and N) | Police, Law Enforcement Funding, Criminal Laws |
| g08 | PR_7 (Y and N) | Renewable Energy Generation |
| g08 | PR_8 (Y and N) | Eliminates Rich of Same-Sex Couples to Marry |
| g08 | PR_9 (Y and N) | Criminal Justice System. Victims' Rights. Parole |
| g08 | PR_2 (Y and N) | Standards for Confining Farm Animals |
| g08 | PR_3 (Y and N) | Children's Hospital Bond Act. Grant Program |
| g08 | PR_4 (Y and N) | Parent Notif. Before Terminating Minor's Pregnancy |

| Election | Office | Rep 1 | Rep 2 |
|---|---|---|---|
| p10 | GOV | Poizner | Whitman |

| Election | Office | Dem 1 | Dem 2 |
|---|---|---|---|
| p10 | LTG | Hahn | Newsom |

| Election | Office | Rep 1 | Rep 2 |
|---|---|---|---|
| p10 | LTG | Maldonado | Aanestad |
| p10 | SOS | Dunn | Taitz |

| Election | Office | Dem 1 | Dem 2 |
|---|---|---|---|
| p10 | INS | Jones | De La Torre |

| Election | Office | Rep 1 | Rep 2 |
|---|---|---|---|
| p10 | INS | Fitzgerald | Villines |
| p10 | CON | Evans | Strickland |

| Election | Office | Dem 1 | Dem 2 |
|---|---|---|---|
| p10 | ATG | Torrico | Harris |

* The specific top-two Democratic and Republican candidates in the State Assembly, Congressional and Senate district races varies by district.

| Election | Office | Rep 1 | Rep 2 |
|---|---|---|---|
| p10 | ATG | Cooley | Eastman |
| p10 | USS | Fiorina | Campbell |

| Election | Proposition | Title |
|---|---|---|
| p10 | PR_13 (Y and N) | Property Taxes and Seismic Retrofit of Buildings |
| p10 | PR_14 (Y and N) | Primary Election Participation |
| p10 | PR_15 (Y and N) | California Fair Elections Act |
| p10 | PR_16 (Y and N) | Local Electricity Providers |
| p10 | PR_17 (Y and N) | Auto Insurance Pricing |

| Election | Office | Dem | Rep |
|---|---|---|---|
| g10 | GOV | Brown | Whitman |
| g10 | LTG | Newsom | Maldonado |
| g10 | SOS | Bowen | Dunn |
| g10 | CON | Chiang | Strickland |
| g10 | INS | Jones | Villines |
| g10 | TRS | Lockyer | Walters |
| g10 | ATG | Harris | Cooley |
| g10 | USS | Boxer | Fiorina |
| g10 | CNG | district candidate* | district candidate* |
| g10 | SEN | district candidate* | district candidate* |
| g10 | ASS | district candidate* | district candidate* |

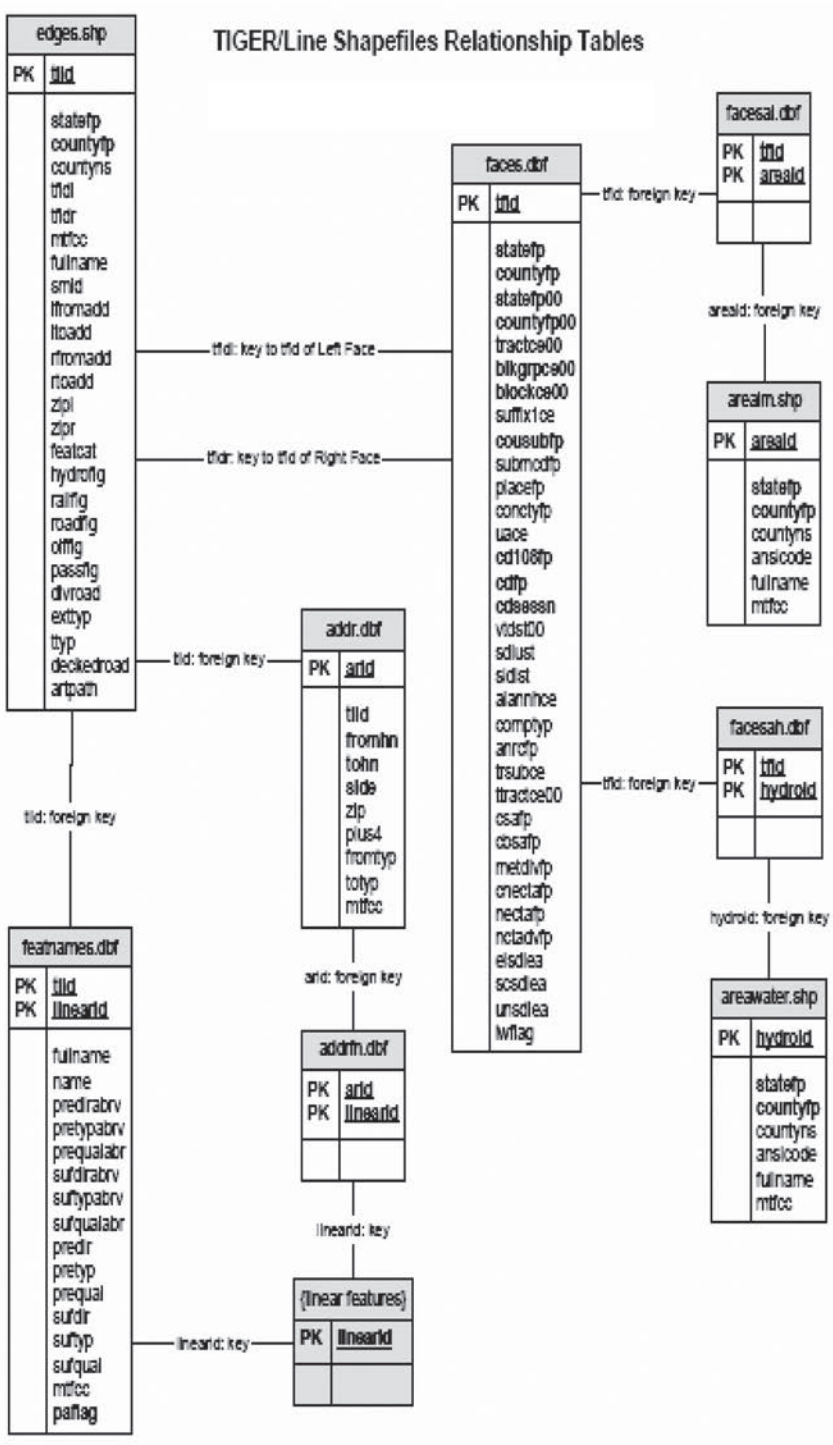
| Election | Proposition | Title |
|---|---|---|
| g10 | PR_19 (Y and N) | Legalize Marijuana in CA, Regulate and Tax |
| g10 | PR_20 (Y and N) | Redistricting of Congressional Districts |
| g10 | PR_21 (Y and N) | State Park Funding. Vehicle License Surcharge |
| g10 | PR_22 (Y and N) | Prohibit State From Taking some Local Funds |
| g10 | PR_23 (Y and N) | Suspend Air Pollution Control Law |
| g10 | PR_24 (Y and N) | Repeal Allowance of Lower Business Tax Liability |
| g10 | PR_25 (Y and N) | Simple Majority Vote to Pass Budget |
| g10 | PR_26 (Y and N) | 2/3 Vote for Some State/Local Fees |
| g10 | PR_27 (Y and N) | Eliminate State Redistricting Commission |

* The specific top-two Democratic and Republican candidates in the State Assembly, Congressional and Senate district races varies by district.

**Appendix IV: TIGER relational data files**

The following documentation is taken from the Census Bureau TIGER file documentation. These are the files which are used to perform the geocoding discussed in this documentation.

***Figure 6: TIGER/LINE Shapefiles Relationship Tables***



The All Lines shapefile (edges.shp) contains the geometry and attributes of each topological primitive edge. Each edge has a unique TLID (TIGER/Line Identifier) value. The left and right faces for an edge can be determined by linking on the TFIDL (left) or TFIDR (right) attribute to the TFID attribute in the Topological Faces relationship table (faces.dbf). The Address Ranges relationship table (addr.dbf) contains the attributes of each address range. Each address range has a unique ARID value. The edge to which an address range applies can be determined by linking to the edges shapefile on the TLID attribute. Multiple address ranges can apply to the same edge (an edge can have multiple address ranges).

The Address Range-Feature Name relationship table (addrfn.dbf) contains a record for each address range-linear feature name relationship. The purpose of this relationship file is to identify all street names associated with each address range. An edge can have several feature names; an address range located on an edge can be associated with one or any combination of the available feature names (an address range can have multiple feature names). The address range is identified by the ARID attribute, which can be used to link to the Address Ranges relationship table. The linear feature name is identified by the LINEARID attribute that relates the address range back to the featnames.dbf table.

The Feature Names relationship table (featnames.dbf) contains a record for each feature name-edge combination, and includes the feature name attributes. The edge to which a Feature Names relationship table record applies can be determined by linking to the All Lines shapefile on the TLID attribute. Multiple Feature Names relationship table records can link to the same edge, for example, a road edge could link to US Hwy 22 and Rathburn Road. The linear feature to which the feature name applies is identified by the LINEARID attribute. Multiple feature names may exist for the same edge (linear features are not included in the data set, but could be constructed using the All Lines shapefile and the relationship tables).

The Topological Faces relationship table contains the attributes of each topological primitive face. Each face has a unique TFID value. The face geometries can be built from the All Lines shapefile using the edges' left and right face relationships. The geometries of each geographic entity can then be built by dissolving the face geometries on the appropriate attribute(s) in the Topological Faces relationship table.

The Area Landmark shapefile (arealm.shp) contains the geometry and attributes of each area landmark. Each area landmark has a unique AREAID value.

The Topological Faces-Area Landmark relationship table (facesal.dbf) contains a record for each face-area landmark relationship. The face to which a Topological Faces-Area Landmark relationship table record applies can be determined by linking to the Topological Faces relationship table on the TFID attribute. The area landmark to which a Topological Faces-Area Landmark relationship table record applies can be determined by linking to the Area Landmark shapefile on the AREAID attribute. A face may be part of multiple area landmarks. An area landmark may consist of multiple faces.

The Area Hydrography shapefile contains the geometry and attributes of each area hydrography feature. Each area hydrography feature has a unique HYDROID value.

The Topological Faces-Area Hydrography relationship table contains a record for each face-area hydrography feature relationship. The face to which a Topological Faces-Area Hydrography relationship table record applies can be determined by linking to the Topological Faces table on the TFID attribute. The area hydrography feature to which a Topological Faces-Area Hydrography relationship table record applies can be determined by linking to the Area Hydrography shapefile on the HYDROID attribute. A face may be part of multiple area water features. An area water feature may consist of multiple faces.

**Appendix V: Ecological techniques**

***What "Ecological" Techniques Do***

Suppose there are two types of voters, left and right. Furthermore, on a proposition (call it Prop X), left voters support X at an 80 percent level and right voters support it at a 20 percent level. Then taking three precincts, we would expect the following:

| **Precinct** | **Prop X Yes** | **Left Voters** | | **Right Voters** | |
|---|---|---|---|---|---|
| | **Vote** | *Left Voters* | *Votes for X* | *Right Voters* | *Votes for X* |
| 1A | 100 | 80 | 64 | 180 | 36 |
| 2A | 120 | 100 | 80 | 200 | 40 |
| 4B | 280 | 320 | 256 | 120 | 24 |

So what do ecological inference techniques do? Note that in the above table, we know the precinct, we know the Prop X Yes Vote (from the SOV), we know the number of left voters (from the county registrar who voted file), and the number of right voters (same source). What we don't know is the level of support of left voters for Prop X, nor the level of support of right voters for Prop X. Ecological inference find the levels of support from election results when those levels of support are not known.

Here's a simple example of how ecological inference works. Suppose we have two precincts only and want to find the level of support of the left and right voters. If it is precinct 1A and 2A, then rewrite the above table for these precincts as

| **Precinct** | **Prop X Yes** | **Left Voters** | | **Right Voters** | |
|---|---|---|---|---|---|
| | **Vote** | *Left Voters* | *Votes for X* | *Right Voters* | *Votes for X* |
| 1A | 100 | 80 | x1 | 180 | x2 |
| 2A | 120 | 100 | x1 | 200 | x2 |

Here, we are uncertain as to the levels of support so we write them as x1 for the support of left voters and x2 as the support of right voters. Then this is two equations in two unknowns (just like in Algebra I), so we solve

100 = 80 * x1 + 180 * x2
120 = 100 * x1 + 200 * x2

for x1 and x2. The solution? x1 = .8 and x2 = .2. For more than two equations, one technique for handling multiple equations is a statistical technique called least squares or regression. Goodman in 1951 was the first to propose this, and it is sometimes call a Goodman regression.

Why are there issues with ecological inference? Well, here is one potential problem. Taking the first table above, let's alter the last row so that instead of 280 votes yes on Prop X, we get a table as follows:

| **Precinct** | **Prop X Yes** | **Left Voters** | | **Right Voters** | |
|---|---|---|---|---|---|
| | **Vote** | *Left Voters* | *Votes for X* | *Right Voters* | *Votes for X* |
| 1A | 100 | 80 | x1 | 180 | x2 |
| 2A | 120 | 100 | x1 | 200 | x2 |
| 4B | 200 | 320 | x1 | 120 | x2 |

Now, suppose we solve the second and third equation for x1 and x2, or the system of equations

120 = 100 * x1 + 200 * x2
200 = 320 * x1 + 120 * x2

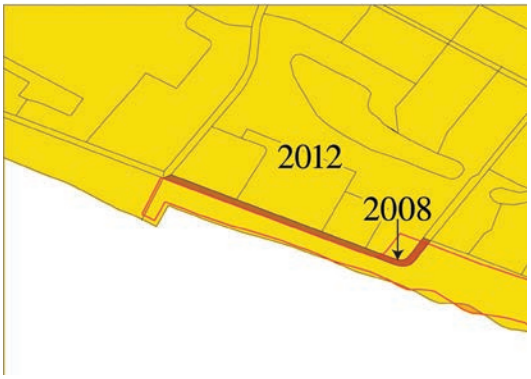
The solution is x1 = .48 and x2 = .36--quite a bit different than the levels of support of .8 and .2.
So a key assumption of ecological inference is that the support levels are (relatively) constant across precincts. Some error is inevitable, and handling this error is typically how variants of ecological regression differ.

While ecological regression is used in this database construction to improve the accuracy of database information, it should be noted that ecological regression is the accepted statistical methodology in Voting Rights Act cases for inferring the propensity of groups of individuals to support political candidates. One court described the method this way: "Ecological regression, the standard method for inferring the behavior of population groups from data collected for aggregate units, was used to estimate the voting behavior of non-Hispanics and Hispanics. The regression methodology generates prediction equations that indicate how voting responds to variations in the proportions of Hispanics and non-Hispanics in each precinct. These equations can provide the information needed to estimate the average voting of non- Hispanics and Hispanics, respectively, in the election district under analysis." Garza v. County of Los Angeles, 756 F. Supp. 1298 (C.D. Cal. 1990), aff'd, 918 F.2d 763, cert. denied, 498 U.S. 1028 (1991).

**Appendix VI: Potential Geocoding Problems**

For an illustration of potential geocoding problems, consider *Figure 2* in the text, reproduced here.

***Figure 2: 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 PL94-171, 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 on the TIGER files do not agree with the Census Bureau's internal assignment files (which are not released). So the primary source of block registration/population errors (zero population blocks with registration, blocks with population and no registration) come about from the mismatches that are described here.