



**Technical Documentation:**  
**California Health Disadvantage Index (HDI 1.1)**

by

Neil Maizlish, PhD, Epidemiologist

Public Health Alliance of Southern California

Email: [admin@PHASoCal.org](mailto:admin@PHASoCal.org)

Tel: (619) 452-1180

Website: <http://phasocal.org/ca-hdi/>

January 26, 2016

## TABLE OF CONTENTS

<b>BACKGROUND</b> .....	1
<b>METHODS/WORKFLOW</b> .....	1
<b>Creating Reference Files</b> .....	2
<b>Evaluating Missing Data</b> .....	3
<b>Imputing Missing Data and Constructing the HDI Score</b> .....	5
<b>DATA PROCESSING PROCEDURES</b> .....	6
<b>DATA QUALITY PROCEDURES</b> .....	8
<b>REFERENCES</b> .....	9
<b>Table 1.</b> Definition of HDI indicators by Domain .....	10
<b>Table 2.</b> Data Dictionary and Source Data Variable Transformations for HDI 1.1 Files .....	11
<b>Table 3.</b> Variables in B17024 to Calculate Proportion of the Population under age 64 with Household Incomes Below Twice the Federal Poverty Line .....	20
<b>Table 4.</b> Reference File Data Dictionary (hdi_shortname.csv) .....	22
<b>Table 5.</b> Example of Standardized HDI Indicator Reference File (hdi_costburden.csv) .....	23
<b>Table 6.</b> Number and Percent of Census Tracts with Missing Data .....	24
<b>Table 7.</b> Data Dictionary for HDI Master Output File (HDI1.1Data2016-01-17.xls) .....	25
<b>Appendix A.</b> Download Instructions for U.S. 2010 Decennial Census and American Community Survey .....	30
<b>Appendix B.</b> Download Instructions for Statewide Election Data.....	34

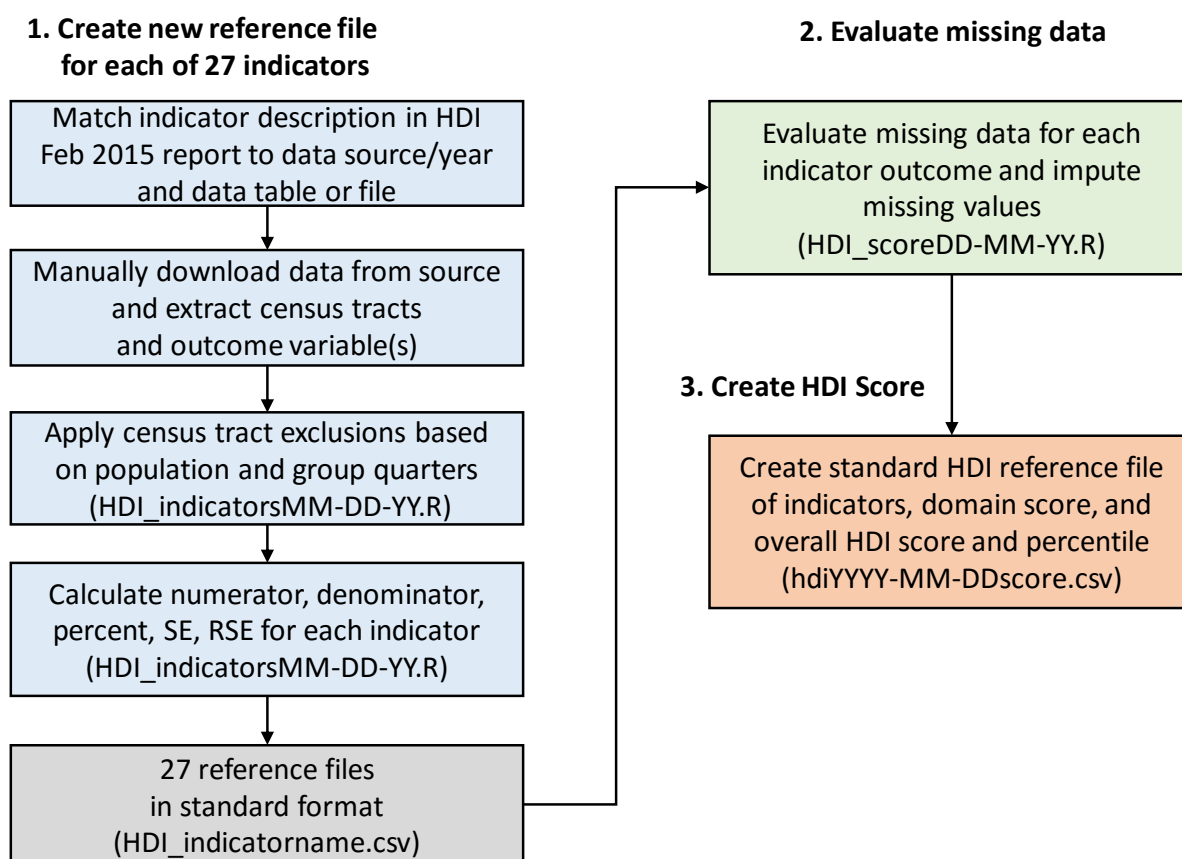
## BACKGROUND

The purpose of this report is to provide technical information on the construction of the California Health Disadvantage Index, HDI, (version 1.1) and its 27 constituent indicators. The background on the development of the HDI is provided in a separate report,<sup>1</sup> which should be read before proceeding. The technical documentation reviews source data and data quality procedures, and provides step-by-step procedures for obtaining source data files and automating data processing using R scripts that read, transform, and reformat source data into standard HDI data files and the HDI score.

## METHODS/WORKFLOW

An overview of the work flow from HDI concept to statistical measure is presented in Figure 1.

### Figure 1. Work Flow of HDI File Creation



The work flow is organized into three phases: 1) creating reference files for each of the 27 indicators that comprise the HDI, 2) an assessment of data quality, especially missing data, and 3) creation of the overall HDI score and a standardized file used in further statistical analyses.

## Creating Reference Files

### Scoping and Variable Matching

In the first phase, the descriptions of the indicator (Table 1) were matched to putative data sources in a narrowing scoping process. Websites, technical manuals, and data dictionaries of the source files were reviewed for specific variables that corresponded to the definitions in Table 1. Source data and specific files tended to be either disaggregated data or pre-aggregated data tables. For many indicators, specific denominator and numerator variables were identified to create a proportion or a rate. To be a match, all aspects of the definition had to be satisfied in the source data, including age ranges and other qualifying criteria.

Tables 2 and 3 presents the 27 HDI indicators (in alphabetic order), the data source name, data source files and specific tables, variable names in the source data, and the URL where source data can be downloaded from the internet. In addition to the numerator and denominator of proportions (percentages), rates, and ratios, Table 2 also presents the source variables names and the formulae for constructing standard errors and relative standard errors of HDI indicators.

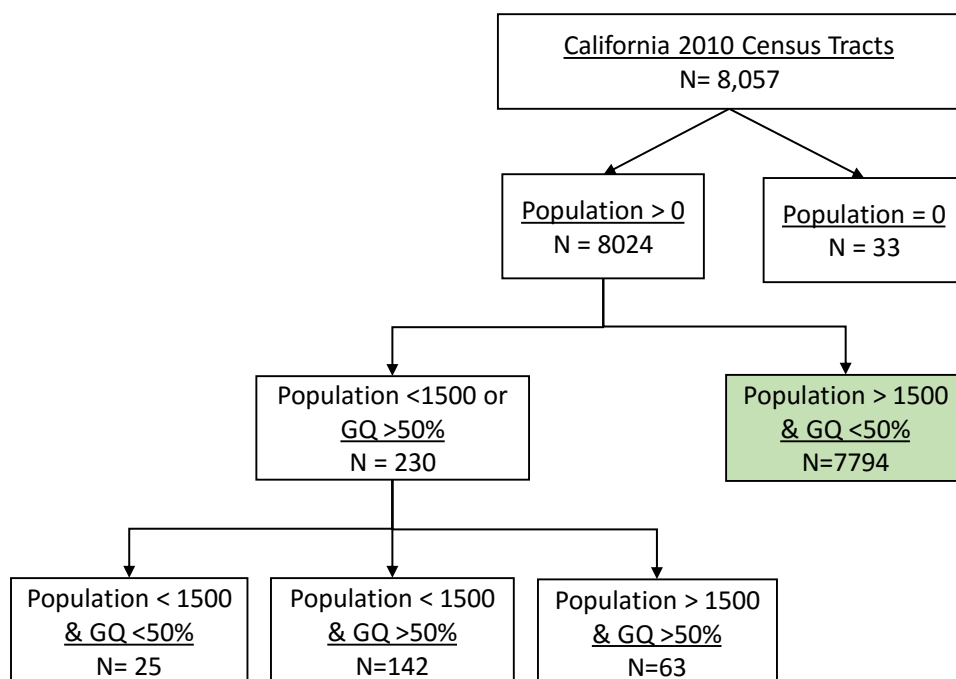
### Data Acquisition

Data were acquired from publically accessible websites (Table 2) except for years of life lost (yll), which was acquired through a request to Dr. Stephen Wolfe, Center on Society and Health, Virginia Commonwealth University. Fifteen indicators derived from the U.S. 2010 Decennial Census<sup>2</sup> and the American Community Survey (2008-2012)<sup>3</sup> were obtained through the American Fact Finder data download interface. Step-by-step instructions are provided in Appendix A for using this interface to select specific data tables. Appendix B provides step-by-step instructions for downloading data on voter participation from the University of California, Berkeley Statewide Database.<sup>4</sup> Other variables were available from Excel spreadsheets or comma separated value (CSV) files. Four variables (asthma, PM2.5, traffic, low birth weight) were in the CalEnviroScreen 2.0<sup>5</sup> Excel spreadsheet produced by California Environmental Protection Agency (CalEPA) from primary data sources. Three indicators (pedshurt, treecanopy, parks) were in spreadsheets produced by the California Department of Public Health<sup>6</sup> from their analysis of primary sources. Spreadsheets containing the retail density indicator were obtained from the U.S. EPA's Smart Location Database.<sup>7</sup> Spreadsheet data on supermarket access were obtained from the Economic Research Service<sup>8</sup> of the U.S. Department of Agriculture, which analyzed primary data.

### Exclusion Criteria and Creating Standard HDI Indicator Files

Census tracts with a population less than 1500 people in 2010 or a group quarters population in 2010 that was greater than or equal to 50 percent of the total census tract population were excluded from analysis (Figure 2). The former criteria aimed to avoid statistically unstable sample estimates in the American Community Survey and the second criteria aimed to preserve the validity of indicators that could produce spurious results by including prisoners, college students, or institutionalized populations who are economically dependent or mobility-limited.

Figure 2. Exclusions for Population and Group Quarters (GQ)



The exclusions were performed as part of R program (version 3.2.1), which also created the indicator values, standard, errors, and relative standard errors for each census tract.

An R program (HDI\_indicatorsMM-DD-YY.R) was written to input source data files manually downloaded from the Internet, extract and transform source data variables into HDI indicators, and output 27 standardize data files. The basic program blocks are presented in Figure 3. The data dictionary for a standard file is presented in Table 4 and an example of the standard file is presented in Table 5. All indicators were operationalized so that a higher value indicated a higher degree of disadvantage. This required that the proportion of the population with park access to be reassigned as lack of park access (1-p) and income and retail were transformed by multiplying their values by -1.

### Evaluating Missing Data

The 27 files generated by the R program HDI\_indicatorsMM-DD-YY.R were input to a second R program (HDI\_scoreMM-DD-YYY.R), which read each file and computed the number and percent of census tracts with missing data in the source file (Figure 4). Table 6 gives descriptive statistics of missing data.\* Approximately six percent of eligible census tracts had missing data for 1 or more missing variables, although the vast majority of census tracts with missing data (395/451, 88%) were missing only one variable.

\* In HDI 1.0, an indicator of proximity to frequent transit was proposed for the neighborhood domain. Because this indicator had a substantial amount missing data in several California communities, an indicator of tree canopy coverage was substituted in the neighborhood domain in HDI 1.1.

Figure 3. Schematic of R Program to Create Indicator Files

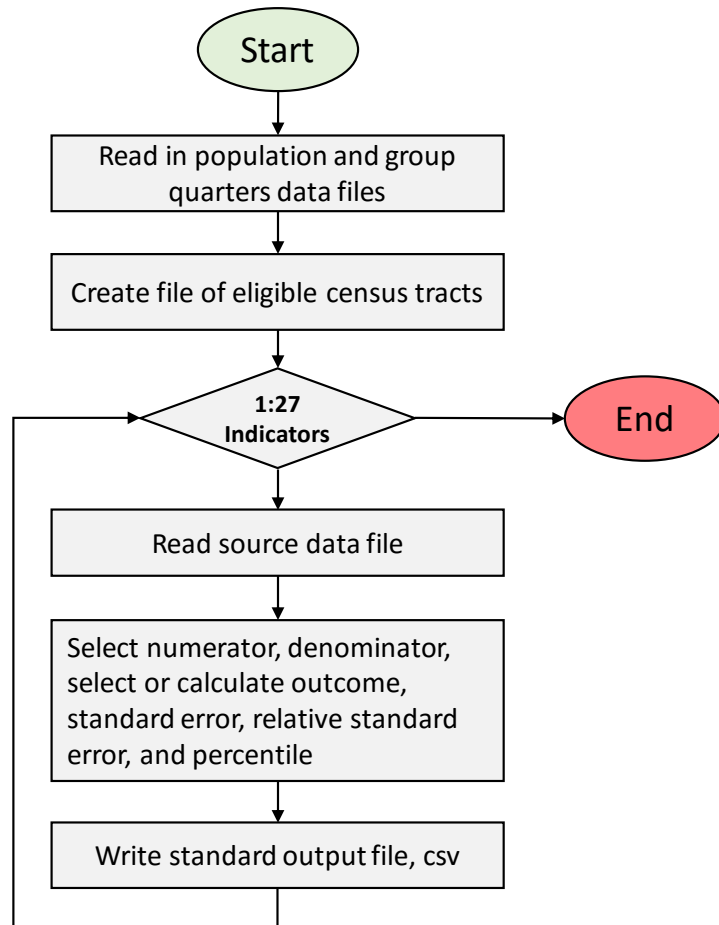
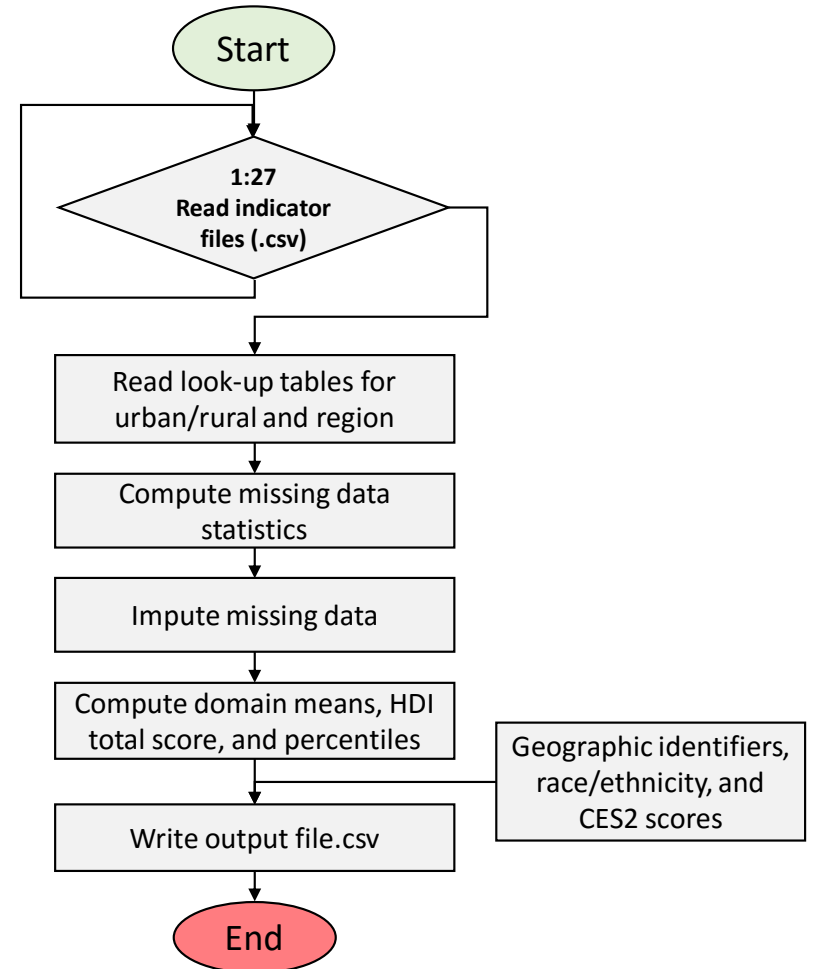


Figure 4. Schematic of R Program to Assess Missing Data and Create HDI Total Score



The Los Angeles County census tract 6037930401, which was erroneously assigned in the 2010 Decennial Census,<sup>9</sup> had 15 missing variables and was excluded in further analyses. Years of life lost had the highest number of census tracts with missing data (2.7%). Missing data for preschool enrollment and years of life lost account for nearly all of rest of missing data. Because the Educational Opportunity domain only has these 2 variables, there is less "buffering" of the effect of missing data. Although the number of impacted census tracts is small, the Educational Opportunity domain mean would be entirely missing for 199 census tracts because there are no other variables in this domain. Rural census tracts and those of the Central Coast, Northern California, and the Sierra region ("Other region") were disproportionately missing data (Table 6).

### Imputing Missing Data and Constructing the HDI Score

Missing data are not collected, or are excluded because errors in data collection produce implausible values, or data suppression rules are imposed to protect confidentiality or exclude statistically unreliable data. When an index is created from multiple variables, missing data in any of the constituent variables will potentially impact the index. Not accounting for missing data may introduce bias into the index.

There are three approaches to deal with missing data in statistical analyses. First is to consider only complete cases. This approach would result in throwing out data on 451 census tracts (Table 6). Second is to use all available data. This would use all non-missing data but lead to a different number of variables factored into the HDI for each census tract. Because the Education domain has only two variables, 199 census tracts are missing for both variables (Table 6) and would not have any data to contribute to the domain mean. Third is to calculate an imputed value. There are many imputation methods, several of which, while simple to implement, may introduce their own bias. For example, substituting missing data with the "mean" will drive the overall mean towards null and reduce the standard deviation (which is used to compute Z scores for the HDI). Single and multiple imputation methods use regression models of non-missing variables to create less biased estimates of missing data. The availability of software that implements multiple imputation makes this option appealing, and, for the HDI, a multiple imputation algorithm for R (Amelia) was used to impute missing data.<sup>10</sup>

After imputing data, the R program (HDI\_scoreMM-DD-YY.R) calculated constrained Z scores for each indicator and domain mean Z scores, and the percentile ranking of each census tract.

The output of the imputation program was a data file with complete data on 27 variables for each of 7,793 eligible census tracts. Z scores of each of the 27 variables were computed for each census tract. For a given variable, the Z score,  $Z$ , for the  $i$ th census tract is the difference between the census tract value,  $X$ , and the overall variable mean,  $\mu$ , divided by the variable's standard deviation,  $\sigma$ . Z scores were constrained to a scale of 0 (least) to 5 (most disadvantaged):

$$Z_i = \frac{X_i - \mu}{\sigma}, \text{ where } 0 \leq Z \leq 5$$

$$Z_i = 0, \text{ where } Z \leq 0$$

$$Z_i = 5, \text{ where } Z \geq 5$$

In the calculation of Z scores for median annual income and retail density, the variable values were initially multiplied by -1 to express smaller incomes and lower retail densities as more disadvantaged (less negative). The resulting Z scores were subsequently multiplied by -1 to re-express disadvantage on the 0 to 5 scale.

The arithmetic mean of Z scores,  $\bar{Z}$ , of variables in each of the 6 domains were computed for each census tract. The HDI for each census tract was computed as the weighted average of the domain means, using the following weights:

$$\text{HDI} = (0.5 \times \bar{Z}_{\text{Economic}}) + (0.1 \times \bar{Z}_{\text{Education}}) + (0.1 \times \bar{Z}_{\text{Environment}}) + (0.1 \times \bar{Z}_{\text{Health}}) + (0.05 \times \bar{Z}_{\text{Neighborhood}}) + (0.15 \times \bar{Z}_{\text{Social}})$$

The census tract percentile of individual variables, domain mean Z scores, and the HDI was based on their rank order among 7793 census tracts. Ties were assigned the arithmetic average of their ranks.

A "master file" of results includes information on all (N=8057) California 2010 census tracts. Including total population, percent living in group quarters, geographic identification information (census tract, city associated with the centroid of the census tract, zip code, county), variable values and their percentile ranking, domain mean Z scores and their percentile ranking, and the HDI score and its percentile ranking. A data dictionary is presented in Table 7. Additional computed fields include race/ethnicity for 8 mutually exclusive major 2010 Census groups (Latino/Hispanic of any race, Black, Asian, American Indian/Alaskan Native, Asian, Native Hawaiian and Other Pacific Islander, Other Race, Multiple races, and Whites), the top 25% of HDI disadvantaged census tracts, quintiles of the HDI distribution, CalEnviroScreen 2.0 percentile, top 25% percentile of CES2.0 disadvantaged census tracts, agreement and disagreement between HDI and CES2 top 25% disadvantaged census tracts. File names and data sources for these additional fields are listed at the end of Table 7.

## DATA PROCESSING PROCEDURES

The overview of file organization for data processing by R programs is presented in Figure 5.

### Inputs

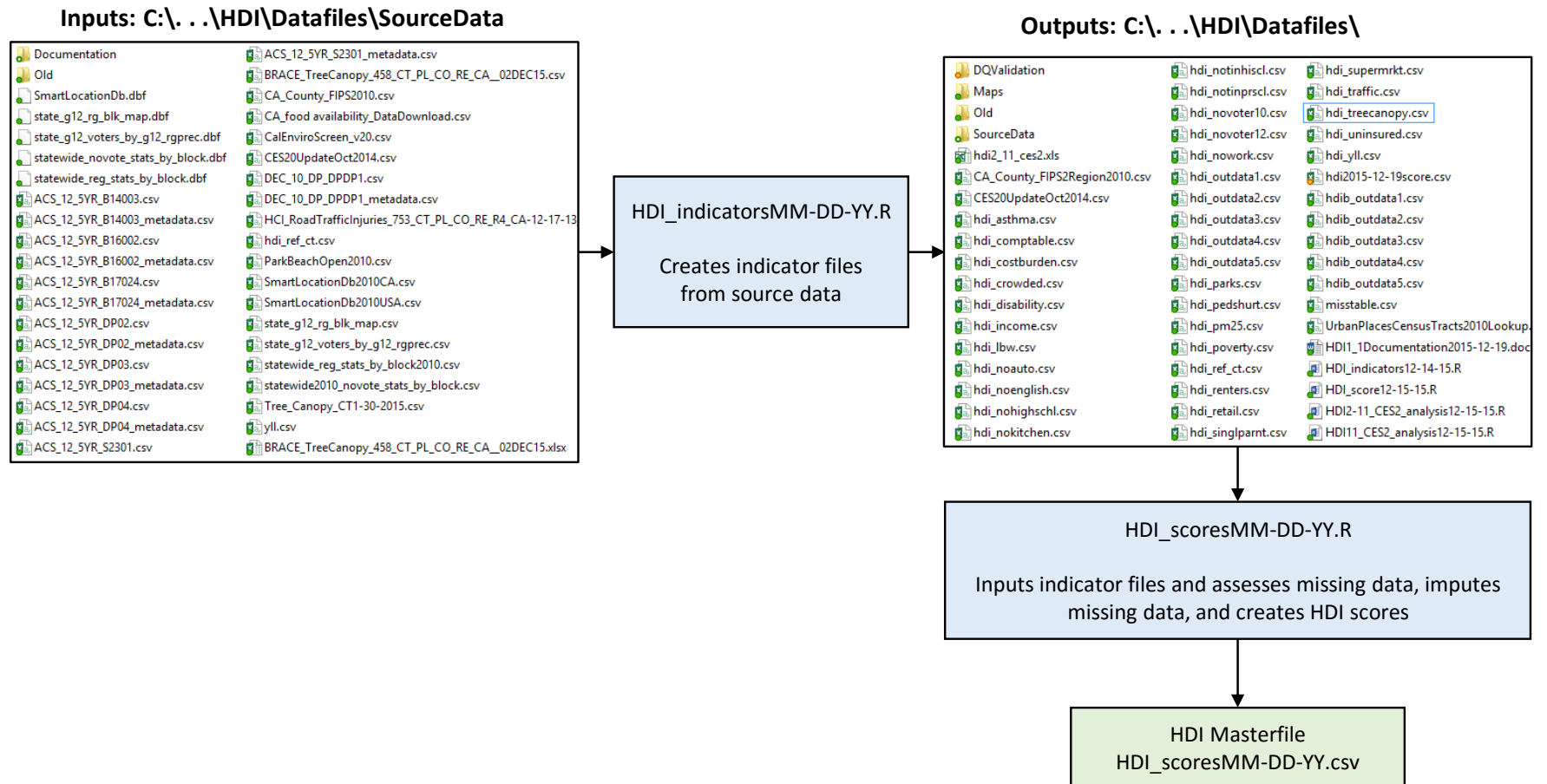
Source data files must be placed in a folder named "C:\. . . \DataFiles\SourceData". (The C:\. . . is part of the path that is specific to a user's computer.) These files include CalEnviroScreen October 2014 Update indicators (asthma, low birth weight, PM2.5, traffic), years of life lost, and downloads from American Community Survey, US EPA's Smart Location Database (retail), and CDPH's Healthy Community Indicators Project (pedestrian injuries, tree canopy, and parks), UC Berkeley Statewide voting database (2010 and 2012 elections), and USDA's food availability (supermarket access). Any source file in an Excel format (xls or xlsx) or dBASE4 (.dbf) format must be saved as a .csv file.

### Outputs

The R Program (HDI\_IndicatorsMM-DD-YY.R) processes the source data for each HDI indicator and creates a separate output file in a standard format (Tables 4-5). The files by default are written to the "C:\. . . \DataFiles\SourceData" folder. These files have a "hdi\_ . . . csv" prefix and extension. Users must check these files for data quality and then manually copy all 27 and a stub file for geo-referencing (hdi\_ref\_ct.csv) to "C:\. . . \DataFiles\" for further analysis with the R program that imputes missing data and creates the HDI score (HDI\_ScoreMM-DD-YY.R).



## Figure 5. Inputs, Outputs, and R Programs for Data Processing



## Imputation

A multiple imputation algorithm for R was used to impute missing data.<sup>10</sup> The following intermediate files were generated by the Amelia R package (called within HDI\_ScoreMM-DD-YY.R) during the multiple imputation iterations: hdi\_outdata1.csv, hdi\_outdata2.csv, hdi\_outdata3.csv, hdi\_outdata4.csv, hdi\_outdata5.csv, hdib\_outdata1.csv, hdib\_outdata2.csv, hdib\_outdata3.csv, and hdib\_outdata4.csv. The file 'hdib\_outdata5.csv' is the final file with imputed values for HDI1.1, and was used in further data processing within the R program.

## **DATA QUALITY PROCEDURES**

### **R Programs and Outputs**

Programming and debugging followed a top-down incremental approach, stopping after each new program block to check output in the R Studio Console and that data transformations occurred properly. The head and tail rows of each intermediate file was checked by manual inspection, and row counts were checked before and after merging of files to verify that the anticipated number of rows was correct. Key output files were manually checked to determine that the following were calculated correctly:

- 1) percents of indicator outcomes in R program output were checked against original data sources, mostly ACS factfinder output (Appendix A)
- 2) Z scores and constrained z scores were manually checked in Excel (z\_score\_hdi\_score\_manual\_validation11-20-15.docx)
- 3) In R output files the direction of disadvantage was checked to verify that a higher percent tracked a higher Z score (manually in Excel output), income and retail flipped (e.g., hdi2015-11-21pct\_z\_tile.xlsx)
- 4) domain means from R program output were manually spot checked from calculations in Excel
- 5) overall weighted HDI scores were manually calculated in Excel and checked against R program output (e.g., hdi2015-11-21pct\_z\_tile.xlsx).

## REFERENCES

1. Bhatia R, Maizlish N. *Methodology for a Health Disadvantage Index for California. Final Report*. Berkeley, CA and Los Angeles, CA: The Civic Engine and the Public Health Alliance of Southern California; 2016.
2. U.S. Census Bureau. *U.S. Decennial Census*. DS. Washington, DC: U.S. Census Bureau; 2010. <http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml>.
3. U.S. Census Bureau. *American Community Survey*. DS. Washington, DC: U.S. Census Bureau; 2008-2012. <http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml>.
4. Berkeley Law Center for Research. *Statewide Database*. Berkeley, CA: University of California; 2015. <http://statewidedatabase.org/index.html>.
5. Faust J, August L, Alexeeff G, Bangia K, Cendak R, Cheung-Sutton E, et al. *California Communities Environmental Health Screening Tool, Version 2.0 (CalEnviroScreen 2.0): Guidance and Screening Tool*. Sacramento, CA: Office of Environmental Health Hazard Assessment, California Environmental Protection Agency; 2014. <http://oehha.ca.gov/ej/pdf/CES20FinalReportUpdateOct2014.pdf>.
6. Office of Health Equity. *Healthy Communities Data and Indicators Project*. Sacramento, CA: California Department of Public Health; 2015. <https://www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx>.
7. Ramsey K, Bell A. *Smart Location Database. Version 2.0 User Guide. Updated: March 14, 2014*. Washington, DC: U.S. Environmental Protection Agency; 2014. [http://www.epa.gov/sites/production/files/2014-03/documents/sld\\_userguide.pdf](http://www.epa.gov/sites/production/files/2014-03/documents/sld_userguide.pdf).
8. Economic Research Service. *Food Environment Atlas (8/19/2015)*. Washington, DC: U.S. Department of Agriculture; 2015. [http://www.ers.usda.gov/datafiles/Food Environment Atlas/Data Access and Documentation on Downloads/Current Version/documentation.pdf](http://www.ers.usda.gov/datafiles/Food%20Environment%20Atlas/Data%20Access%20and%20Documentation%20on%20Downloads/Current%20Version/documentation.pdf).
9. U.S. Census Bureau. *2010 Census Geography Note 5, California*. Washington, DC: U.S. Bureau of Census; 2013. [http://www2.census.gov/geo/pdfs/reference/Geography\\_Notes.pdf](http://www2.census.gov/geo/pdfs/reference/Geography_Notes.pdf).
10. Honaker J, King G, Blackwell M. Amelia II: A Program for Missing Data. *J Statistical Software* 2011;45:1-47. <https://cran.r-project.org/web/packages/Amelia/vignettes/amelia.pdf>.

**Tables**

**Table 1.** Definition of HDI indicators by Domain

Domain	Short Name	Definition
Economic	costburden	Percentage of renter households paying more than 30% of income on rent
	crowded	Percentage of households with more than 1 occupant per room
	income	Median household income
	noauto	Percentage of households without access to an automobile
	nokitchen	Percentage of the population in homes lacking complete kitchen facilities
	nowork	Percentage of population aged 25-64 who are unemployed
	poverty	Percentage of the population under aged 65 with household incomes below twice the Federal Poverty Line
	uninsured	Percentage of the population without health insurance
Education	notinhschl	Percentage of 15-17 year olds not enrolled in school
	notinprsch	Percentage of 3 and 4 year olds not enrolled in school
Environment	pedshurt	Annual rate of pedestrian injuries
Environment	pm25	Annual average PM 2.5 level
Environment	traffic	Traffic density on highways within 150 feet of census tract boundaries
Health	asthma	Annual rate of emergency room visits for asthma
	disability	Percentage of the non-institutionalized population with any disability
	lbw	Percentage of live born infants with low birth weights
	yll	Years of life lost per capita
Neighborhood	parks	Percentage of the population not living within a half-mile of a park, beach, or open space greater than 1 acre
	retail	Combined employment density for retail, entertainment, and educational uses (jobs/acre)
	supermrkt	Percentage of the population living more than one mile from a supermarket or large grocery store
	treecanopy	Population-weighted percentage of the census tract area without tree canopy
Social	noenglishathome	Percentage of household where no person at least 14 years old speaks English well
	nohighschl	Percentage of population over age 25 without a high school education
	novoter10	Percentage of registered voters not voting in the 2010 general election
	novoter12	Percentage of registered voters not voting in the 2012 general election
	renters	Percentage of occupied housing units not occupied by property owners
	singlparnt	Percentage of family households with children under 18 with only one parent

**Table 2.** Data Dictionary and Source Data Variable Transformations for HDI 1.1 Files

Variable Name	Data Source	Table	Variable(s)
Total Population	2010 Census	DEC_10_DPDP1	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) GEO.ID2 (FIPS), HD01_S001 (Total Population) See Appendix A for download instructions
Group Quarters	2010 Census	DEC_10_DPDP1	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) Group Quarters Population (HD01_S143); Group Quarters Percent (HD02_S143)
asthma	CES2	CES2OCT2014	URL: <a href="http://oehha.ca.gov/ej/pdf/CES20UpdateOct2014.xlsx">http://oehha.ca.gov/ej/pdf/CES20UpdateOct2014.xlsx</a> Asthma (Note: change Census Tract to FIPS)
costburden	ACS5YR2012	DP04	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) HC01_VC191 Estimate; GROSS RENT AS A%AGE OF HOUSEHOLD INCOME - Occupied units paying rent (excluding units where GRAPI cannot be computed) HC01_VC196 Estimate; GROSS RENT AS A%AGE OF HOUSEHOLD INCOME - 30.0 to 34.9% HC04_VC196 Percent MOE; GROSS RENT AS A%AGE OF HOUSEHOLD INCOME - 30.0 to 34.9% HC01_VC197 Estimate; GROSS RENT AS A%AGE OF HOUSEHOLD INCOME - 35.0% or more HC04_VC197 Percent MOE; GROSS RENT AS A%AGE OF HOUSEHOLD INCOME - 35.0% or more Numerator = HC01_VC196 + HC01_VC197 Denominator = HC01_VC191 Percent = Numerator/denominator $se\_pct = [(HC04\_VC196/1.645)^2 + (HC03\_VC197/1.645)^2]^{0.5}$ $rse = se\_pct / Percent$ , if percent <.5 and $rse = se\_pct / (1-Percent)$ , if percent >.5 (if percent = 0, the rse = 100 very large value)
crowded	ACS5YR2012	DP04	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) HC01_VC109 Estimate; OCCUPANTS PER ROOM - Occupied housing units HC01_VC111 Estimate; OCCUPANTS PER ROOM - 1.01 to 1.50 HC04_VC111 Percent Margin of Error; OCCUPANTS PER ROOM - 1.01 to 1.50 HC01_VC112 Estimate; OCCUPANTS PER ROOM - 1.51 or more HC04_VC112 Percent Margin of Error; OCCUPANTS PER ROOM - 1.51 or more Numerator = HC01_VC111 + HC01_VC112 Denominator = HC01_VC109 Percent = Numerator/denominator $se\_pct = [(HC04\_VC111/1.645)^2 + (HC03\_VC112/1.645)^2]^{0.5}$

Variable Name	Data Source	Table	Variable(s)
disability	ACS5YR2012	DP02	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) HC01_VC103 Estimate; Total Civilian Noninstitutionalized Population HC01_VC104 Estimate; With a disability HC04_VC104 Percent Margin of Error; With a disability Numerator = HC01_VC104 Denominator = HC01_VC103 Percent = Numerator/denominator se_pct = HC04_VC104/1.645
income	ACS5YR2012	DP03	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) HC01_VC74 Estimate; INCOME AND BENEFITS (IN 2012 INFLATION-ADJUSTED DOLLARS) - Total households HC01_VC85 Estimate; INCOME AND BENEFITS (IN 2012 INFLATION-ADJUSTED DOLLARS) - Median household income (dollars) HC02_VC85 Margin of Error; INCOME AND BENEFITS (IN 2012 INFLATION-ADJUSTED DOLLARS) - Median household income (dollars)  Denominator = HC01_VC74 Numerator = HC01_VC85 se_pct = HC02_VC85/1.645
lbw	CES2	CES2OC T2014	URL: <a href="http://oehha.ca.gov/ej/pdf/CES20UpdateOct2014.xlsx">http://oehha.ca.gov/ej/pdf/CES20UpdateOct2014.xlsx</a> Low Birth Weight
noauto	ACS5YR2012	DP04	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) HC01_VC81 Estimate; VEHICLES AVAILABLE - Occupied housing units HC01_VC82 Estimate; VEHICLES AVAILABLE - No vehicles available HC04_VC82 Percent Margin of Error; VEHICLES AVAILABLE - No vehicles available Denominator = HC01_VC81 Numerator = HC01_VC82 se_pct = HC04_VC82/1.645

Variable Name	Data Source	Table	Variable(s)
noenglishathome	ACS5YR2012	B16002	<p>URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download)</p> <p>HD01_VD01 Estimate; Total: Households            HD02_VD01 Margin of Error; Total: households            HD01_VD04 Estimate; Total: - Spanish: - No one 14 and over speaks English only or speaks English "very well"            HD02_VD04 Margin of Error; Total: - Spanish: - No one 14 and over speaks English only or speaks English "very well"            HD01_VD07 Estimate; Total: - Other Indo-European languages: - No one 14 and over speaks English only or speaks English "very well"            HD02_VD07 Margin of Error; Total: - Other Indo-European languages: - No one 14 and over speaks English only or speaks English "very well"            HD01_VD10 Estimate; Total: - Asian and Pacific Island languages: - No one 14 and over speaks English only or speaks English "very well"            HD02_VD10 Margin of Error; Total: - Asian and Pacific Island languages: - No one 14 and over speaks English only or speaks English "very well"            HD01_VD13 Estimate; Total: - Other languages: - No one 14 and over speaks English only or speaks English "very well"            HD02_VD13 Margin of Error; Total: - Other languages: - No one 14 and over speaks English only or speaks English "very well"            Denominator = HD01_VD01            Numerator = HD01_VD04 + HD01_VD07 + HD01_VD10 + HD01_VD13            SE(numerator, X) = sqrt((HD02_VD04/1.645)<sup>2</sup> + (HD02_VD07/1.645)<sup>2</sup> + (HD02_VD10/1.645)<sup>2</sup> + (HD02_VD13/1.645)<sup>2</sup>)            SE(denominator, Y) = HD02_VD01/1.645</p> $SE(\hat{\rho}) = \frac{1}{\hat{\rho}} \sqrt{[SE(\hat{X})]^2 - \frac{\hat{X}^2}{\hat{\rho}^2} [SE(\hat{Y})]^2}$

Variable Name	Data Source	Table	Variable(s)
nohighschl	ACS5YR2012	DP02	<p>URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download)</p> <p>HC01_VC84 Estimate; EDUCATIONAL ATTAINMENT - Population 25 years and over            HC01_VC85 Estimate; EDUCATIONAL ATTAINMENT - Less than 9th grade            HC01_VC86 Estimate; EDUCATIONAL ATTAINMENT - 9th to 12th grade, no diploma            HC04_VC93 Percent Margin of Error; EDUCATIONAL ATTAINMENT - Percent high school graduate or higher</p> <p>Denominator = HC01_VC84            Numerator = HC01_VC85 + HC01_VC86            se_pct = HC04_VC93/1.645            Note se_pct is same for p and 1-p</p>
nokitchen	ACS5YR2012	DP04	<p>URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download)</p> <p>HC01_VC102 Estimate; SELECTED CHARACTERISTICS - Occupied housing units            HC01_VC104 Estimate; SELECTED CHARACTERISTICS - Lacking complete kitchen facilities            HC04_VC104 Percent Margin of Error; SELECTED CHARACTERISTICS - Lacking complete kitchen facilities</p> <p>Denominator = HC01_VC102            Numerator = HC01_VC104            se_pct =( HC04_VC104/1.645)/100</p>
notinhiscl	ACS5YR2012	B14003	<p>URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download)</p> <p>HD01_VD07 Estimate; Male: - Enrolled in public school: - 15 to 17 years            HD01_VD16 Estimate; Male: - Enrolled in private school: - 15 to 17 years            HD01_VD25 Estimate; Male: - Not enrolled in school: - 15 to 17 years            HD01_VD35 Estimate; Female: - Enrolled in public school: - 15 to 17 years            HD01_VD44 Estimate; Female: - Enrolled in private school: - 15 to 17 years            HD01_VD53 Estimate; Female: - Not enrolled in school: - 15 to 17 years</p> <p>Denominator = HD01_VD07 + HD01_VD16 + HD01_VD25 + HD01_VD35 + HD01_VD44 + HD01_VD53            Numerator = HD01_VD25 + HD01_VD53</p> <p>HD02_VD07 Margin of Error; Male: - Enrolled in public school: - 15 to 17 years            HD02_VD16 Margin of Error; Male: - Enrolled in private school: - 15 to 17 years            HD02_VD25 Margin of Error; Male: - Not enrolled in school: - 15 to 17 years            HD02_VD35 Margin of Error; Female: - Enrolled in public school: - 15 to 17 years            HD02_VD44 Margin of Error; Female: - Enrolled in private school: - 15 to 17 years            HD02_VD53 Margin of Error; Female: - Not enrolled in school: - 15 to 17 years</p> <p><math>SE(\text{denominator}, Y) = \sqrt{(HD02\_VD07/1.645)^2 + (HD02\_VD16/1.645)^2 + (HD02\_VD25/1.645)^2 + (HD02\_VD35/1.645)^2 + (HD02\_VD44/1.645)^2 + (HD02\_VD53/1.645)^2}</math>  <math>SE(\text{numerator}, X) = \sqrt{(HD02\_VD25/1.645)^2 + (HD02\_VD53/1.645)^2}</math></p>



Variable Name	Data Source	Table	Variable(s)
notinprsl	ACS5YR2012	B14003	<p style="text-align: center;"><math>SE(\hat{P}) = \frac{1}{\hat{Y}} \sqrt{[SE(\hat{X})]^2 - \frac{\hat{X}^2}{\hat{Y}^2} [SE(\hat{Y})]^2}</math></p> <p>URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download)</p> <p>HD01_VD04 Estimate; Male: - Enrolled in public school: - 3 and 4 years            HD01_VD13 Estimate; Male: - Enrolled in private school: - 3 and 4 years            HD01_VD22 Estimate; Male: - Not enrolled in school: - 3 and 4 years            HD01_VD32 Estimate; Female: - Enrolled in public school: - 3 and 4 years            HD01_VD41 Estimate; Female: - Enrolled in private school: - 3 and 4 years            HD01_VD50 Estimate; Female: - Not enrolled in school: - 3 and 4 years</p> <p>Denominator = HD01_VD04 + HD01_VD13 + HD01_VD22 + HD01_VD32 + HD01_VD41 + HD01_VD50            Numerator = HD01_VD22 + HD01_VD50</p> <p>HD02_VD04 Margin of Error; Male: - Enrolled in public school: - 3 and 4 years            HD02_VD13 Margin of Error; Male: - Enrolled in private school: - 3 and 4 years            HD02_VD22 Margin of Error; Male: - Not enrolled in school: - 3 and 4 years            HD02_VD32 Margin of Error; Female: - Enrolled in public school: - 3 and 4 years            HD02_VD41 Margin of Error; Female: - Enrolled in private school: - 3 and 4 years            HD02_VD50 Margin of Error; Female: - Not enrolled in school: - 3 and 4 years</p> <p>SE(denominator, Y) = sqrt( (HD02_VD04/1.645)<sup>2</sup> + (HD02_VD13/1.645)<sup>2</sup> + (HD02_VD22/1.645)<sup>2</sup> + (HD02_VD32/1.645)<sup>2</sup> + (HD02_VD41/1.645)<sup>2</sup> + (HD02_VD50/1.645)<sup>2</sup> )            SE(numerator, X) = sqrt( (HD02_VD22/1.645)<sup>2</sup> + (HD02_VD50/1.645)<sup>2</sup> )</p> <p style="text-align: center;"><math>SE(\hat{P}) = \frac{1}{\hat{Y}} \sqrt{[SE(\hat{X})]^2 - \frac{\hat{X}^2}{\hat{Y}^2} [SE(\hat{Y})]^2}</math></p>

Variable Name	Data Source	Table	Variable(s)
novoter10	UCB	See Appendix B	<p>URL: <a href="http://statewidedatabase.org/d10/g10_registration.html">http://statewidedatabase.org/d10/g10_registration.html</a>  statewide_novote_stats_by_block.dbf , statewide_reg_stats_by_block.dbf  R program (HDI_voter_reg10-28-15.R) aggregated these block level files to census tract and produced a raw output file (hdi_vr2010.csv ) with registered voters and nonvoters  Numerator = totreg_n  Denominator = totreg_r  Percent = numerator/denominator; se_pct = sqrt(percent*(1-percent)/denominator)</p>
novoter12	UCB	See Appendix B	<p>URL: <a href="http://statewidedatabase.org/d10/g12_geo_conv.html">http://statewidedatabase.org/d10/g12_geo_conv.html</a>  Registration crosswalk file (state_g12_rg_blk_map.csv) was right-joined on RGPREC_KEY to voter file (state_g12_voters_by_g12_rgprec.csv).</p> <p>denominator = BLKREG  numerator = TOTREG_R (voters)  percent = 1-numerator/denominator  se_pct = se_pct = sqrt(percent*(1-percent)/denominator)</p> <p>Note: because the allocation of precinct to block is imperfect some parts of precincts were not allocated to blocks (precinct registration = 17,981,054 vs. block registration = 17,840,280; difference = 140,774, 1%)</p> <p>Method of analysis suggested by Janine Heiser, UCB (510) 624-9086</p>
nowork	ACS2008-12	S2301	<p>URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download)  Note: the age range is specific for 25 to 64 years, which is found in a 2 way table with educational attainment, using total marginals  HC01_EST_VC38 Total; Estimate; EDUCATIONAL ATTAINMENT - Population 25 to 64 years  HC02_EST_VC38 In labor force; Estimate; EDUCATIONAL ATTAINMENT - Population 25 to 64 years, %  HC04_EST_VC38 Unemployment rate; Estimate; EDUCATIONAL ATTAINMENT - Population 25 to 64 years  HC04_MOE_VC38 Unemployment rate; Margin of Error; EDUCATIONAL ATTAINMENT - Population 25 to 64 years  Denominator = HC01_EST_VC38 * HC02_EST_VC38/100  Numerator = (HC04_EST_VC38/100)*Denominator  Percent = Numerator/denominator or HC04_EST_VC38/100  se_pct=(HC04_MOE_VC38/1.645)/100</p>

Variable Name	Data Source	Table	Variable(s)
parks	CaLANDS	HCI	Download ParkBeachOpen10_output4-12-13.zip from <a href="https://www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx">https://www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx</a> Filter Excel file for CT and Total Race/ethnicity, copy and save as .csv file Denominator = pop2010 Numerator = pop_park_acc Percent = 1- numerator/denominator se_pct = se
pedshurt	SWITRS	HCI	Note this is the 5-year (2006-2010) annual average rate of severe and fatal pedestrian injuries per 100,000 population Download two files (HCI_RoadTrafficInjuries_753_CT_PL_CO_RE_R4_CA-12-17-13_A-N.zip, HCI_RoadTrafficInjuries_753_CT_PL_CO_12-17-13_O-Y.zip) from <a href="https://www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx">https://www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx</a> Filter for reportyear (2006-2010), geotype (CT), and mode (pedestrian), copy and paste as .csv then add A-N and O-Y counties into single file (HCI_RoadTrafficInjuriesA-Z.csv) Serious and fatal injuries and rates are added together. R Program does this aggregation  se_pct = sqrt(Numerator)/Denominator  Note: census tracts with no recorded injuries in SWITRS are set to 0 injuries and 0 injury rate
pm25	CES2	CES20 CT2014	URL: <a href="http://oehha.ca.gov/ej/pdf/CES20UpdateOct2014.xlsx">http://oehha.ca.gov/ej/pdf/CES20UpdateOct2014.xlsx</a> PM2.5 changed to PM25 (Note: change Census Tract to FIPS)
poverty	ACS5YR2012	B17024	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) See Table 3 for variables
renters	ACS5YR2012	DP04	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) HC01_VC62 Estimate; HOUSING TENURE - Occupied housing units HC01_VC64 Estimate; HOUSING TENURE - Renter-occupied HC04_VC64 Percent Margin of Error; HOUSING TENURE - Renter-occupied  Denominator = HC01_VC62 Numerator = HC01_VC64 Percent = numerator/denominator se_pct =(HC04_VC64/1.645)/100

Variable Name	Data Source	Table	Variable(s)
retail	USEPA	SmartLocationData base 2.0 (2013)	<a href="http://www2.epa.gov/smartgrowth/smart-location-mapping#SLD">http://www2.epa.gov/smartgrowth/smart-location-mapping#SLD</a> Note: files is at census block group (CBG) and must be aggregated to census tract D1C8_Ret10 Gross retail (8-tier) employment density (jobs/acre) on unprotected land D1C8_Ent10 Gross entertainment (8-tier) employment density (jobs/acre) on unprotected land D1C8_Ed10 Gross education(8-tier) employment density (jobs/acre) on unprotected land Ac_Unpr Total land area in acres that is not protected from development (i.e., not a park or conservation area)  Percent = $D1c8\_Ret10 + D1c8\_Ent10 + D1c8\_Ed10$ Denominator = $Ac\_Unpr * Denominator$ Numerator = percent*
singlparnt	ACS5YR2012	DP02	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) HC01_VC06 Estimate; HOUSEHOLDS BY TYPE - Family households (families) - With own children under 18 years HC01_VC10 Estimate; HOUSEHOLDS BY TYPE - Family households (families) - Male householder, no wife present, family - With own children under 18 years HC04_VC10 Percent Margin of Error; HOUSEHOLDS BY TYPE - Family households (families) - Male householder, no wife present, family - With own children under 18 years HC01_VC12 Estimate; HOUSEHOLDS BY TYPE - Family households (families) - Female householder, no husband present, family - With own children under 18 years HC04_VC12 Percent Margin of Error; HOUSEHOLDS BY TYPE - Family households (families) - Female householder, no husband present, family - With own children under 18 years  Denominator = $HC01\_VC06$ Numerator = $HC01\_VC10 + HC01\_VC12$ Percent = numerator/denominator $se\_pct = \sqrt{((HC04\_VC10/1.645)^2 + (HC04\_VC12/1.645)^2)}/100$
supermrkt	USDA	Food Access Research Atlas	<a href="http://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data.aspx">http://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data.aspx</a> filter Excel file for California, and save as .csv (CA_food availability_DataDownload.csv) CensusTract = FIPS numerator = lapop1 denominator = POP2010 percent = numerator/denominator (=lapop1share) $se\_pct = \sqrt{\text{percent} * (1 - \text{percent}) / \text{denominator}}$
traffic	CES2	CES2OCT2014	URL: <a href="http://oehha.ca.gov/ej/pdf/CES20UpdateOct2014.xlsx">http://oehha.ca.gov/ej/pdf/CES20UpdateOct2014.xlsx</a> Traffic

Variable Name	Data Source	Table	Variable(s)
---------------	-------------	-------	-------------

treecanopy	HCI		Download BRACE_TreeCanopy_458_CT_PL_CO_RE_CA__02DEC15.xlsx from: <a href="https://www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx">https://www.cdph.ca.gov/programs/Pages/HealthyCommunityIndicators.aspx</a> Filter for race_eth_name (Total), geotype (CT), and strata_level_name (population-weighted), copy and paste as .csv percent = percent se_pct =percent_se rse_pct = percent_rse
uninsured	ACS5YR2012	DP03	URL: <a href="http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml">http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml</a> (download) HC01_VC127 Estimate; HEALTH INSURANCE COVERAGE - Civilian noninstitutionalized population HC01_VC131 Estimate; HEALTH INSURANCE COVERAGE - No health insurance coverage HC04_VC131 Percent Margin of Error; HEALTH INSURANCE COVERAGE - No health insurance coverage Denominator = HC01_VC127 Numerator = HC01_VC131 Percent = numerator/denominator se_pct =(HC04_VC131/1.645)/100
yll	VCU	HDI_FEB 2015	yll.csv (provided directly from Virginia Commonwealth University, October 2014) yll = percent

**Table 3.** Variables in B17024 to Calculate Proportion of the Population under age 64 with Household Incomes Below Twice the Federal Poverty Line

<p>Denominator HD01_VD02    Estimate; Under 6 years:</p> <p>Numerator HD01_VD03    Estimate; Under 6 years: - Under .50 HD01_VD04    Estimate; Under 6 years: - .50 to .74 HD01_VD05    Estimate; Under 6 years: - .75 to .99 HD01_VD06    Estimate; Under 6 years: - 1.00 to 1.24 HD01_VD07    Estimate; Under 6 years: - 1.25 to 1.49 HD01_VD08    Estimate; Under 6 years: - 1.50 to 1.74 HD01_VD09    Estimate; Under 6 years: - 1.75 to 1.84 HD01_VD10    Estimate; Under 6 years: - 1.85 to 1.99</p>	<p>Denominator HD01_VD54    Estimate; 25 to 34 years:</p> <p>Numerator HD01_VD55    Estimate; 25 to 34 years: - Under .50 HD01_VD56    Estimate; 25 to 34 years: - .50 to .74 HD01_VD57    Estimate; 25 to 34 years: - .75 to .99 HD01_VD58    Estimate; 25 to 34 years: - 1.00 to 1.24 HD01_VD59    Estimate; 25 to 34 years: - 1.25 to 1.49 HD01_VD60    Estimate; 25 to 34 years: - 1.50 to 1.74 HD01_VD61    Estimate; 25 to 34 years: - 1.75 to 1.84 HD01_VD62    Estimate; 25 to 34 years: - 1.85 to 1.99</p>
<p>Denominator HD01_VD15    Estimate; 6 to 11 years:</p> <p>Numerator HD01_VD16    Estimate; 6 to 11 years: - Under .50 HD01_VD17    Estimate; 6 to 11 years: - .50 to .74 HD01_VD18    Estimate; 6 to 11 years: - .75 to .99 HD01_VD19    Estimate; 6 to 11 years: - 1.00 to 1.24 HD01_VD20    Estimate; 6 to 11 years: - 1.25 to 1.49 HD01_VD21    Estimate; 6 to 11 years: - 1.50 to 1.74 HD01_VD22    Estimate; 6 to 11 years: - 1.75 to 1.84 HD01_VD23    Estimate; 6 to 11 years: - 1.85 to 1.99</p>	<p>Denominator HD01_VD67    Estimate; 35 to 44 years:</p> <p>Numerator HD01_VD68    Estimate; 35 to 44 years: - Under .50 HD01_VD69    Estimate; 35 to 44 years: - .50 to .74 HD01_VD70    Estimate; 35 to 44 years: - .75 to .99 HD01_VD71    Estimate; 35 to 44 years: - 1.00 to 1.24 HD01_VD72    Estimate; 35 to 44 years: - 1.25 to 1.49 HD01_VD73    Estimate; 35 to 44 years: - 1.50 to 1.74 HD01_VD74    Estimate; 35 to 44 years: - 1.75 to 1.84 HD01_VD75    Estimate; 35 to 44 years: - 1.85 to 1.99</p>
<p>Denominator HD01_VD28    Estimate; 12 to 17 years:</p> <p>Numerator HD01_VD29    Estimate; 12 to 17 years: - Under .50 HD01_VD30    Estimate; 12 to 17 years: - .50 to .74 HD01_VD31    Estimate; 12 to 17 years: - .75 to .99 HD01_VD32    Estimate; 12 to 17 years: - 1.00 to 1.24 HD01_VD33    Estimate; 12 to 17 years: - 1.25 to 1.49 HD01_VD34    Estimate; 12 to 17 years: - 1.50 to 1.74 HD01_VD35    Estimate; 12 to 17 years: - 1.75 to 1.84 HD01_VD36    Estimate; 12 to 17 years: - 1.85 to 1.99</p>	<p>Denominator HD01_VD80    Estimate; 45 to 54 years:</p> <p>Numerator HD01_VD81    Estimate; 45 to 54 years: - Under .50 HD01_VD82    Estimate; 45 to 54 years: - .50 to .74 HD01_VD83    Estimate; 45 to 54 years: - .75 to .99 HD01_VD84    Estimate; 45 to 54 years: - 1.00 to 1.24 HD01_VD85    Estimate; 45 to 54 years: - 1.25 to 1.49 HD01_VD86    Estimate; 45 to 54 years: - 1.50 to 1.74 HD01_VD87    Estimate; 45 to 54 years: - 1.75 to 1.84 HD01_VD88    Estimate; 45 to 54 years: - 1.85 to 1.99</p>
<p>Denominator HD01_VD41    Estimate; 18 to 24 years:</p>	<p>Denominator HD01_VD93    Estimate; 55 to 64 years:</p>

Numerator		Numerator	
HD01_VD42	Estimate; 18 to 24 years: - Under .50	HD01_VD94	Estimate; 55 to 64 years: - Under .50
HD01_VD43	Estimate; 18 to 24 years: - .50 to .74	HD01_VD95	Estimate; 55 to 64 years: - .50 to .74
HD01_VD44	Estimate; 18 to 24 years: - .75 to .99	HD01_VD96	Estimate; 55 to 64 years: - .75 to .99
HD01_VD45	Estimate; 18 to 24 years: - 1.00 to 1.24	HD01_VD97	Estimate; 55 to 64 years: - 1.00 to 1.24
HD01_VD46	Estimate; 18 to 24 years: - 1.25 to 1.49	HD01_VD98	Estimate; 55 to 64 years: - 1.25 to 1.49
HD01_VD47	Estimate; 18 to 24 years: - 1.50 to 1.74	HD01_VD99	Estimate; 55 to 64 years: - 1.50 to 1.74
HD01_VD48	Estimate; 18 to 24 years: - 1.75 to 1.84	HD01_VD100	Estimate; 55 to 64 years: - 1.75 to 1.84
HD01_VD49	Estimate; 18 to 24 years: - 1.85 to 1.99	HD01_VD101	Estimate; 55 to 64 years: - 1.85 to 1.99

**Table 4.** Reference File Data Dictionary (hdi\_shortname.csv)

Variable	Definition	Format
indicator	Description of Indicator	text
reportyear	Year indicator data is reported	yyyy-yyyy
FIPS	Census tract ID	10 digit ID number (state-county-tract)
County_Name	Name of county in which census tract resides	
County_FIPS	FIPS number for county	4 digit ID number( "6" + 3-digits)
shortname_d	Denominator of indicator	integer
shortname_nr	Numerator of indicator	Integer
shortname_pct	Indicator outcome as percent or a rate	0-100 or rate
shortname_se_pct	Standard error of the percent	0-100 or rate
shortname_rse	Relative standard error of the percent or rate	0-100; 1000 (0 denominator)
shortname_pctile	Percentile distribution of outcome (shortname_pct)	1-100 (least-most disadvantaged)
source	Data source of indicator	Source abbreviation, table or variable name
version	Date/time the file was created	

\* Missing data indicated by "NA"

† "shortname" is the Variable Name in Table 1



**Table 5.** Example of Standardized HDI Indicator Reference File (hdi\_costburden.csv)

indicator	reportyear	FIPS	costburden_n	costburden_d	costburden_pct	costburden_se_pct	costburden_rse	costburden_pctile	source	version
Share of renter households paying more than 30% of income on rent	2008-2012	6001400100	94	240	0.3917	0.1461	0.3731	0.8817	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001400200	79	262	0.3015	0.0956	0.3171	0.9565	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001400300	623	1326	0.4698	0.0733	0.1560	0.7546	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001400400	435	1039	0.4187	0.0652	0.1558	0.8396	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001400500	379	896	0.4230	0.0758	0.1791	0.8328	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001400600	232	416	0.5577	0.1407	0.3182	0.5462	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001400700	696	1167	0.5964	0.0756	0.1874	0.4332	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001400800	547	1006	0.5437	0.0895	0.1962	0.5831	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001400900	301	574	0.5244	0.0911	0.1916	0.6348	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001401000	904	1431	0.6317	0.0758	0.2057	0.3315	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001401100	853	1501	0.5683	0.0732	0.1695	0.5142	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001401200	232	747	0.3106	0.0696	0.2241	0.9520	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001401300	753	1377	0.5468	0.0708	0.1562	0.5745	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001401400	659	1111	0.5932	0.0628	0.1543	0.4420	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001401500	324	495	0.6545	0.0963	0.2789	0.2677	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001401600	378	542	0.6974	0.1093	0.3613	0.1693	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001401700	292	595	0.4908	0.0987	0.2011	0.7111	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001401800	361	514	0.7023	0.0914	0.3072	0.1593	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001402200	402	616	0.6526	0.0895	0.2575	0.2734	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001402400	593	923	0.6425	0.0826	0.2310	0.3012	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001402500	287	569	0.5044	0.0982	0.1981	0.6819	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001402700	258	431	0.5986	0.1261	0.3141	0.4269	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001402800	844	1492	0.5657	0.0566	0.1303	0.5216	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001403000	340	862	0.3944	0.0708	0.1794	0.8767	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001403100	185	327	0.5657	0.1152	0.2654	0.5214	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001403300	882	1621	0.5441	0.0722	0.1583	0.5818	ACS_12_5YR_DP04	Oct 22 2015
Share of renter households paying more than 30% of income on rent	2008-2012	6001403400	1234	2435	0.5068	0.0595	0.1207	0.6746	ACS_12_5YR_DP04	Oct 22 2015

Note: County\_FIPS and County\_name not shown so table can fit to page width

**Table 6.** Number and Percent of Census Tracts with Missing Data

Item	N, Census Tracts	Percent	Item	N, Census Tracts	Percent
<u>HDI Variable</u>			<u>Domain (Number of Variables)</u>		
Asthma	0	0.0	Economic Resources, N=8	5	0.1
Costburden	3	0.0	Social Resources, N=6	14	0.2
Crowded	1	0.0	Educational Opportunity, N=2	199	2.6
Disability	1	0.0	Health Outcomes, N=4	210	2.7
Income	3	0.0	Environmental Hazards, N=3	78	1.0
Noauto	1	0.0	Complete Neighborhoods, N=4	0	0.0
Noenglish	1	0.0			
Nohighschl	1	0.0	<u>Urban/Rural</u>		
Nokitchen	1	0.0	Urban	385	5.1
Notinhiscl	83	1.1	Rural	66	18.1*
Notinprscl	145	1.9			
Nowork	1	0.0	<u>Regions<sup>†</sup></u>		
Poverty	1	0.0	Bay Area	101	6.5*
Renters	1	0.0	Inland Valley	44	5.5
Singlparnt	14	0.2	Los Angeles	155	5.5
Uninsured	1	0.0	Other	71	9.7
Low Birth Weight	2	0.0	Sacramento Area	30	5.9
PM2.5	78	1.0	San Diego	30	4.7
Traffic	0	0.0	San Joaquin Valley	20	2.7
No Voter 2010	0	0.0			
No Voter 2012	0	0.0	<u>Missing Variables/Census Tract</u>		
Parks	0	0.0	0	7343	94.2
Pedshurt	0	0.0	1	395	5.1
Supermrkt	0	0.0	2	38	0.5
Retail	0	0.0	3 -5	17	0.2
Treecanopy	0	0.0	15	1	<0.1
YLL	209	2.7			

\* difference  $p < 0.05$ 

† Regions by County:

Bay Area: Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, SonomaSan Joaquin Valley: Fresno, Kern, Kings, Madera, Merced, San Joaquin, Stanislaus, TulareInland Valley: Riverside, San BernardinoSacramento Area: El Dorado, Placer, Sacramento, Sutter, Yolo, YubaSan Diego: Imperial, San DiegoOther: Butte, Alpine, Amador, Calaveras, Inyo, Mariposa, Mono, Tuolumne, Monterey, San Benito, Santa Cruz, Del Norte, Humboldt, Lake, Mendocino, Trinity, Lassen, Modoc, Nevada, Plumas, Sierra, Siskiyou, Colusa, Glenn, Tehama, San Luis Obispo, Santa Barbara, Shasta, Ventura

**Table 7.** Data Dictionary for HDI Master Output File (HDI1.1Data2016-01-17.xls)

Variable Name	Definition	Code/Comments	Source
CensusTract	10-digit census tract code (state+county+tractID)	leading 0 for state is not included	DEC_10_DP_DPDP1
pop2010	total population of census tract in 2010	Decennial Census 2010	DEC_10_DP_DPDP1
pct2010gq	population in group quarters in 2010	range 0-100	DEC_10_DP_DPDP1
City	City associated with centroid of census tract	spatially computed field	PHASC
ZIP	5-digit postal zip code	USPS	CES2OCT2014
County_FIPS	5-digit code of county	leading 0 for state is not included	DEC_10_DP_DPDP1
County_Name	Name of county	Alameda . . . Yuba	DEC_10_DP_DPDP1
UrbanType	Census classification of urban type	urban, urban_cluster, rural	DEC_10_DP_DPDP1
hdi_total	Total score of HDI	weighted average of domain means	PHASC
hdi_pctile	Percentile of HDI	0-100 (least-most disadvantaged)	PHASC
quintiles	Quintile of HDI score	1st, 2nd, 3rd, 4th, 5th (least-most disadvantaged)	PHASC
quartiles	Quartile of HDI score	1st, 2nd, 3rd, 4th (least-most disadvantaged)	PHASC
hdi_top25pct	Top 25% most disadvantaged	Yes=top 25%, No= least 75%	PHASC
economic	Economic domain mean of 8 constrained z-scores	Highest value= most disadvantaged, 0=least	PHASC
economic_pctile	Percentile of economic domain z score	0-100 (least-most disadvantaged)	PHASC
education	Education domain mean of constrained z-scores	Highest value= most disadvantaged, 0=least	PHASC
educationc_pctile	Percentile of education domain z score	0-100 (least-most disadvantaged)	PHASC
environment	Environment domain mean of constrained z-scores	Highest value= most disadvantaged, 0=least	PHASC
environment_pctile	Percentile of environment domain z score	0-100 (least-most disadvantaged)	PHASC
health	Health domain mean of constrained z-scores	Highest value= most disadvantaged, 0=least	PHASC
health_pctile	Percentile of health domain z score	0-100 (least-most disadvantaged)	PHASC
neighborhood	Neighborhood domain mean of constrained z-scores	Highest value= most disadvantaged, 0=least	PHASC
neighborhood_pctile	Percentile of neighborhood domain z score	0-100 (least-most disadvantaged)	PHASC
social	Economic domain mean of constrained z-scores	Highest value= most disadvantaged, 0=least	PHASC
social_pctile	Percentile of social domain z score	0-100 (least-most disadvantaged)	PHASC

Variable Name	Definition	Code/Comments	Source
asthma	Spatially modeled, age-adjusted rate of emergency department (ED) visits for asthma per 10,000	≥0 (0 least disadvantaged)	CES2OCT2014
asthma_pctile	Percentile of asthma rate	0-100 (least-most disadvantaged)	CES2OCT2014
costburden	Percentage of renter households paying more than 30% of income on rent	0-100 (least-most disadvantaged)	ACS5YR2012_DP04
costburden_pctile	Percentile of percentage of renter cost burden	0-100 (least-most disadvantaged)	PHASC
crowded	Percentage of households with more than 1 occupant per room	0-100 (least-most disadvantaged)	ACS5YR2012_DP04
crowded_pctile	Percentile of crowded households	0-100 (least-most disadvantaged)	PHASC
disability	Percentage of the non-institutionalized population with any disability	0-100 (least-most disadvantaged)	ACS5YR2012_DP02
disability_pctile	Percentile of disability households	0-100 (least-most disadvantaged)	PHASC
income	Median annual household income	≥ 0	ACS5YR2012_DP03
income_pctile	Percentile of median annual income	0-100 (least-most disadvantaged)	PHASC
lbw	Percent low birth weight, spatially modeled	0-100 (least-most disadvantaged)	CES2OCT2014
lbw_pctile	Percentile of percent low birth weight	0-100 (least-most disadvantaged)	PHASC
noauto	Percentage of households without access to an automobile	0-100 (least-most disadvantaged)	ACS5YR2012_DP04
noauto_pctile	Percentile of households without access to an automobile	0-100 (least-most disadvantaged)	PHASC
noenglish	Percentage of household where no person at least 14 years old speaks English well	0-100 (least-most disadvantaged)	ACS5YR2012_B16002
noenglish_pctile	Percentile of household where no person at least 14 years old speaks English well	0-100 (least-most disadvantaged)	PHASC
nohighschl	Percentage of population over age 25 without a high school education	0-100 (least-most disadvantaged)	ACS5YR2012_DP02
nohighschl_pctile	Percentile of population over age 25 without a high school education	0-100 (least-most disadvantaged)	PHASC
nokitchen	Percentage of the population in homes lacking complete kitchen facilities	0-100 (least-most disadvantaged)	ACS5YR2012_DP04
nokitchen_pctile	Percentile of the population in homes lacking complete kitchen facilities	0-100 (least-most disadvantaged)	PHASC
notinhiscl	Percentage of 15-17 year olds not enrolled in school	0-100 (least-most disadvantaged)	ACS5YR2012_B14003
notinhiscl_pctile	Percentile of 15-17 year olds not enrolled in school	0-100 (least-most disadvantaged)	PHASC
notinprsc1	Percentage of 3 and 4 year olds not enrolled in school	0-100 (least-most disadvantaged)	ACS5YR2012_B14003
notinprsc1_pctile	Percentile of 3 and 4 year olds not enrolled in school	0-100 (least-most disadvantaged)	PHASC
Variable Name	Definition	Code/Comments	Source

novoter10	Percentage of registered voters not voting in the 2010 general election	0-100 (least-most disadvantaged)	UCBSWDB2010
novoter10_pctile	Percentile of registered voters not voting in the 2010 general election	0-100 (least-most disadvantaged)	PHASC
novoter12	Percentage of registered voters not voting in the 2012 general election	0-100 (least-most disadvantaged)	UCBSWDB2012
novoter12_pctile	Percentile of registered voters not voting in the 2012 general election	0-100 (least-most disadvantaged)	PHASC
nowork	Percentage of population aged 25-64 who are unemployed	0-100 (least-most disadvantaged)	ACS5YR2012_S2301
nowork_pctile	Percentile of population aged 25-64 who are unemployed	0-100 (least-most disadvantaged)	PHASC
poverty	Percentage of the population under aged 65 with household incomes below twice the Federal Poverty Line	0-100 (least-most disadvantaged)	ACS5YR2012_B17024
poverty_pctile	Percentile of the population under aged 65 with household incomes below twice the Federal Poverty Line	0-100 (least-most disadvantaged)	PHASC
poverty_top25pct	Top 25% of poverty	Yes=top 25%, No= least 75%	PHASC
parks	Percentage of the population not living within a half-mile of a park, beach, or open space greater than 1 acre	0-100 (least-most disadvantaged)	CALANDS2010_HCI
parks_pctile	Percentile of the population not living within a half-mile of a park, beach, or open space greater than 1 acre	0-100 (least-most disadvantaged)	PHASC
pedshurt	Annual rate of pedestrian injuries	0-100 (least-most disadvantaged)	SWITRS2010_HCI
pedshurt_pctile	Percentile of annual rate of pedestrian injuries	0-100 (least-most disadvantaged)	PHASC
pm25	Annual average PM 2.5 level	0-100 (least-most disadvantaged)	CES2OCT2014
pm25_pctile	Percentile of annual average PM 2.5 level	0-100 (least-most disadvantaged)	PHASC
renters	Percentage of occupied housing units not occupied by property owners	0-100 (least-most disadvantaged)	ACS5YR2012_DP04
renters_pctile	Percentile of percentage of occupied housing units not occupied by property owners	0-100 (least-most disadvantaged)	PHASC
retail	Combined employment density for retail, entertainment, and educational uses	0-100 (least-most disadvantaged)	EPASLD2010
retail_pctile	Percentile of employment density for retail, entertainment, and educational uses	0-100 (least-most disadvantaged)	PHASC
singlparnt	Percentage of family households with children under 18 with only one parent	0-100 (least-most disadvantaged)	ACS5YR2012_DP02
singlparnt_pctile	Percentile of percentage of family households with children under 18 with only one parent	0-100 (least-most disadvantaged)	PHASC
supermrkt	Percentage of the population living more than one mile from a supermarket or large grocery store	0-100 (least-most disadvantaged)	USDAERSFARA2010
Variable Name	Definition	Code/Comments	Source

supermrkt_pctile	Percentile of percentage of the population living more than one mile from a supermarket or large grocery store	0-100 (least-most disadvantaged)	PHASC
traffic	Traffic density on highways within 150 feet of census tract boundaries	0-100 (least-most disadvantaged)	CES2OCT2014
traffic_pctile	Percentile of traffic density on highways within 150 feet of census tract boundaries	0-100 (least-most disadvantaged)	PHASC
treecanopy	Population-weighted percentage of the census tract area without tree canopy	0-100 (least-most disadvantaged)	NLMD2011_HCI
treecanopy_pctile	Percentile of population-weighted Percentage of the census tract area without tree canopy	0-100 (least-most disadvantaged)	PHASC
uninsured	Percentage of the population without health insurance	0-100 (least-most disadvantaged)	ACS5YR2012_DP03
uninsured_pctile	Percentile of percentage of the population without health insurance	0-100 (least-most disadvantaged)	PHASC
yll	Years of life lost per capita	0-100 (least-most disadvantaged)	VCU2010
yll_pctile	Percentile of Years of life lost per capita	0-100 (least-most disadvantaged)	PHASC
asian_pct	Percent of Asians in the total population	0-100	
black_pct	Percent of Blacks in the total population	0-100	
latino_pct	Percent of Latinos in the total population	0-100	
multiple_pct	Percent of two or more races in the total population	0-100	
NativeAm_pct	Percent of American Indian/Alaskan Natives in the total population	0-100	
other_pct	Percent of some other race in the total population	0-100	
PacificIsl_pct	Percent of Native Hawaiians and other Pacific Islanders in the total population	0-100	
white_pct	Percent of Whites in the total population	0-100	
CES2Score	CalEnviroScreen 2.0 overall score	Higher is more disadvantaged	CES2OCT2014
CES2PercentileRange	Percentile Range of CES2	(least), 100 (Most disadvantaged)	CES2OCT2014
ces2_pctile	Percentile ranking of CalEnviroScreen 2.0 overall score	0-100 (least-most disadvantaged)	CES2OCT2014
quintiles_ces2	Quintile of CES2 percentile distribution	"1st", "2nd", "3rd", "4th", "5 <sup>th</sup> (most)"	CES2OCT2014
ces2_top25pct	Top 25% of CES2 disadvantaged census tracts	Yes, No	CES2OCT2014
hdi11_ces2	Comparison of HDI1.1 and CES census tracts in the top 25% disadvantaged census tracts	"Agree", "HDI+", "CES-", "HDI-CES+"	PHASC
version	Date file was created by PHASC	Day of Week, Month Day HH:MM:SS YYYY	PHASC

**Table 7. Key for Abbreviations**

ACS, American Community Survey; CES2, CalEnviroScreen2.0 (CalEPA); CALANDS, California Lands Database; DEC, Decennial 2010 Census; EPASLD, Smart Location database (USEPA); HCI, Health Communities Data and Indicators Project (CDPH) NLMD, National Lands Management Database; PHASC, Public Health Alliance of Southern California; SWITRS, Statewide Traffic Records Systems; UCBSWDB, UC Berkeley Statewide Database; VCU, USDAERSFARA, Food Access Research Atlas UCDA/ERS); Virginia Commonwealth University

Look-up Table Files that Associate Census Tract with City, Urban/Rural Areas, Race/Ethnicity, and CalEnviroScreen Scores

Topic	File Name	Data Source
City	census_tracts2_place_look-up_table.csv	HCI/PHASC
Urban/rural	UrbanPlacesCensusTracts2010Lookup.csv	Decennial Census 2010
Race/Ethnicity	pop2010race_eth.csv	DEC_10_SF2_PCT1
CalEnviroScreen 2.0	CES20UpdateOct2014.csv	CalEPA

## Appendix A. Download Instructions for U.S. 2010 Decennial Census and American Community Survey

1. Go to FactFinder Website for Advanced Search (<http://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml>)

U.S. Department of Commerce  
**United States**  
**Census**  
 Bureau

AMERICAN  
**FactFinder**

MAIN | COMMUNITY FACTS | GUIDED SEARCH | **ADVANCED SEARCH** | DOWNLOAD CENTER

**Search** - Use the options on the left (topics, geographies, ...) to narrow your search results

**Your Selections**  
 'Your Selections' is empty  
[load search](#) | [save search](#)

**Search using the options below:**

- Topics** (age, income, year, dataset, ...)
- Geographies** (states, counties, places, ...)
- Race and Ethnic Groups** (race, ancestry, tribe)
- Industry Codes** (NAICS industry, ...)
- EEO Occupation Codes** (executives, analysts, ...)

**To search for tables and other files in American FactFinder:**

- 1 Enter search terms and an optional geography and click GO

topic or table name:  state, county or place (optional):  **GO** ?

topics  race/ancestry  industries  occupations

-- or --

Select from **Topics, Race and Ethnic Groups, Industry Codes, EEO Occupation Codes.**

- these are added to 'Your Selections'
- the Search Results are updated

- 2 Next, select **Geographies** (states, counties, cities, towns, etc.)
  - these are added to 'Your Selections'
  - the Search Results are updated
- 3 Select one or more Search Results and click **View**

2. Pick table: Type table number in dialogue box (ignore autocomplete suggestions)> Go

U.S. Department of Commerce  
**United States**  
**Census**  
 Bureau

AMERICAN  
**FactFinder**

MAIN | COMMUNITY FACTS | GUIDED SEARCH | **ADVANCED SEARCH** | DOWNLOAD CENTER

**Search** - Use the options on the left (topics, geographies, ...) to narrow your search results

**Your Selections**  
 'Your Selections' is empty  
[load search](#) | [save search](#)

**Search using the options below:**

- Topics** (age, income, year, dataset, ...)

**To search for tables and other files in American FactFinder:**

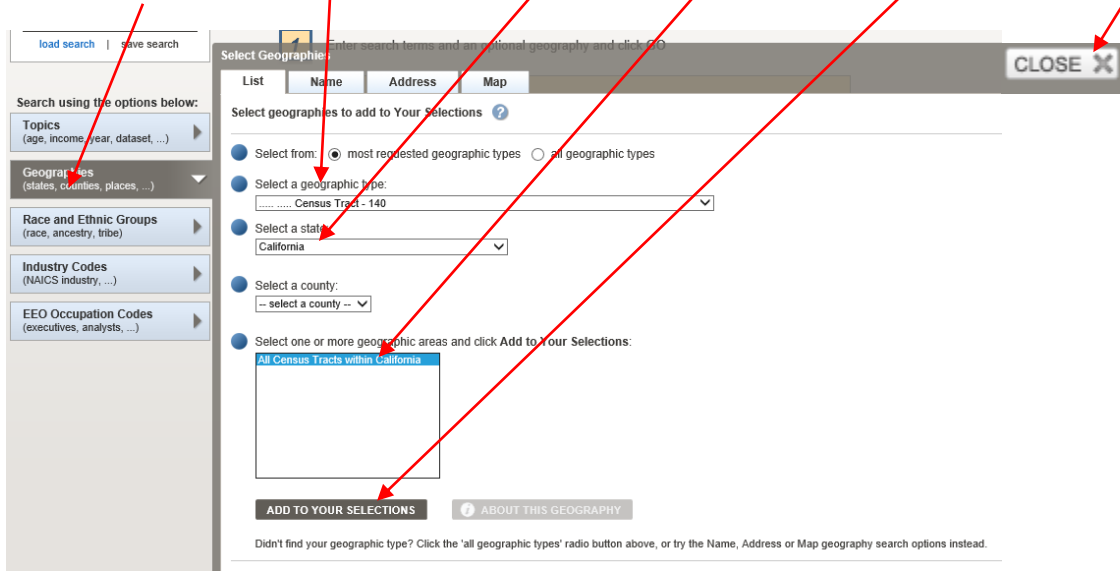
- 1 Enter search terms and an optional geography and click GO

topic or table name:  state, county or place (optional):  **GO** ?

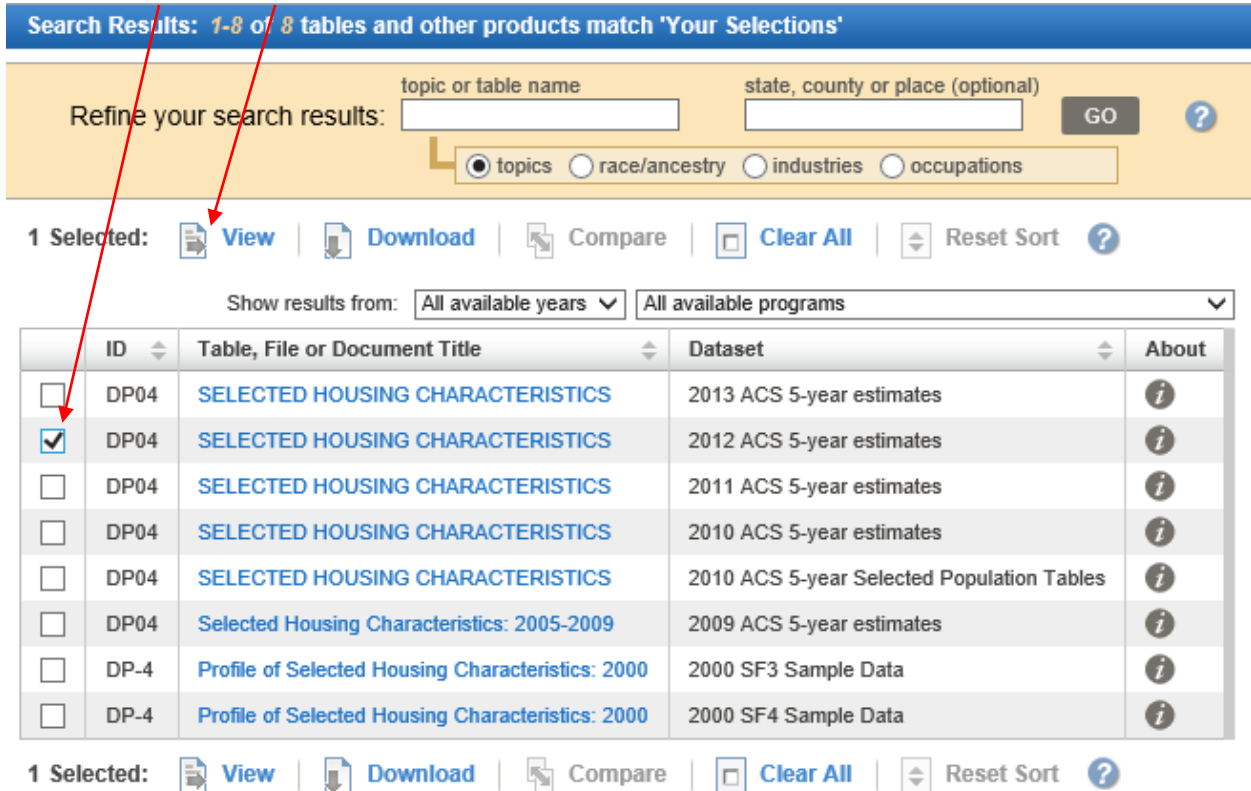
topics  race/ancestry  industries  occupations



3. Pick: Geographies > Census Tract > California > All Census tracks > Add to your Selection > Close

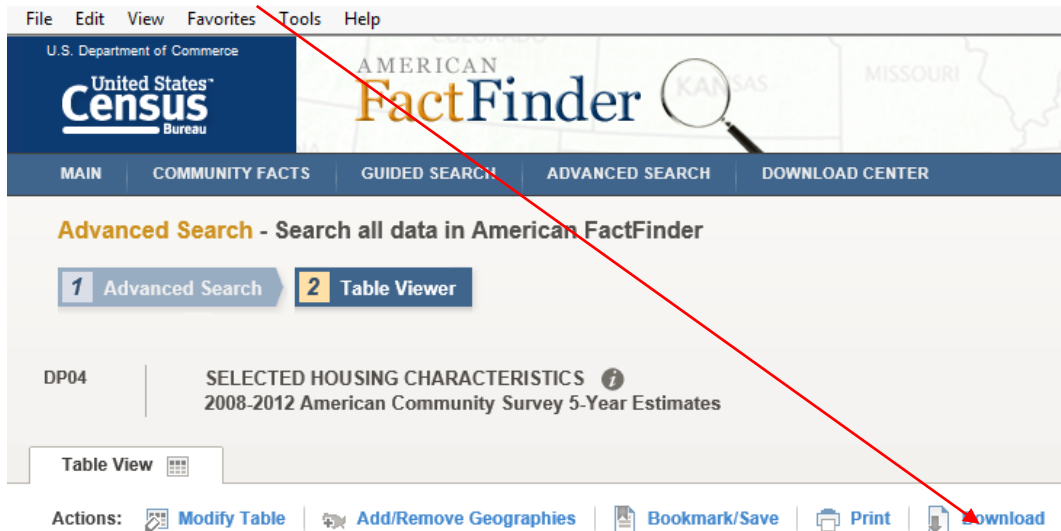


4. Pick Time Period > View



Note: Selecting > View rather than > Download enables a different download screen option for separating the data file from annotations. The resulting data file (e.g., ACS\_12\_5YR\_TABLENUMBER.csv) will only contain numeric information. Files with annotations include symbols that are not numeric (\*, X, etc.) Selected population tables have race/ethnicity detail, but are not used in HDI.

5. Pick Download



Although the American Community Survey (ACS) produces population, demographic and housing unit estimates, it is the Census Bureau's Pop the nation, states, counties, cities and towns and estimates of housing units for states and counties.

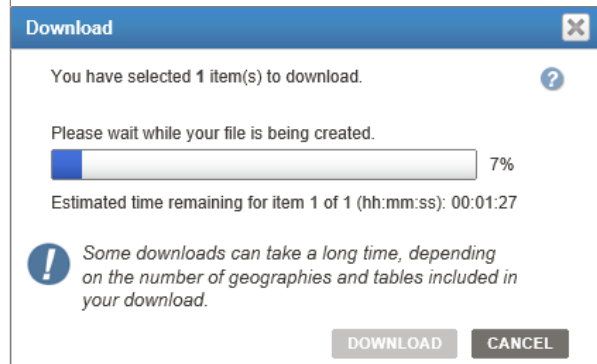
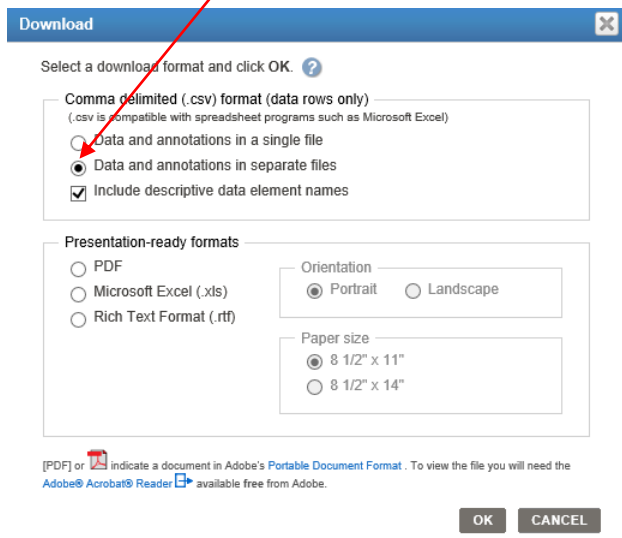
Versions of this table are available for the following years:

- 2013
- 2012
- 2011
- 2010

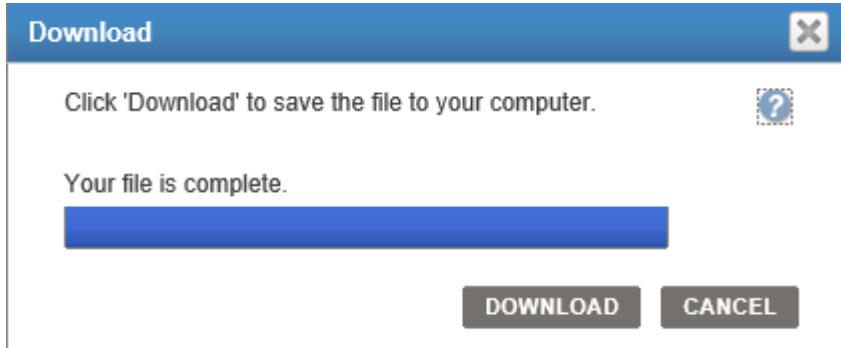
<< 1 - 18 of 32,228 >>

Subject	Census Tract 4001, Alameda County, California				Census Tract 4002, Alameda County, California			
	Estimate	Margin of Error	Percent	Percent Margin of Error	Estimate	Margin of Error	Percent	Percent Margin of Error
HOUSING OCCUPANCY								
Total housing units	1,413	+/-28	1,413	(X)	913	+/-22	913	(X)
Occupied housing units	1,295	+/-72	91.6%	+/-4.5	867	+/-46	95.0%	+/-4.5

6. Annotations in separate file > OK > Wait for download to be created



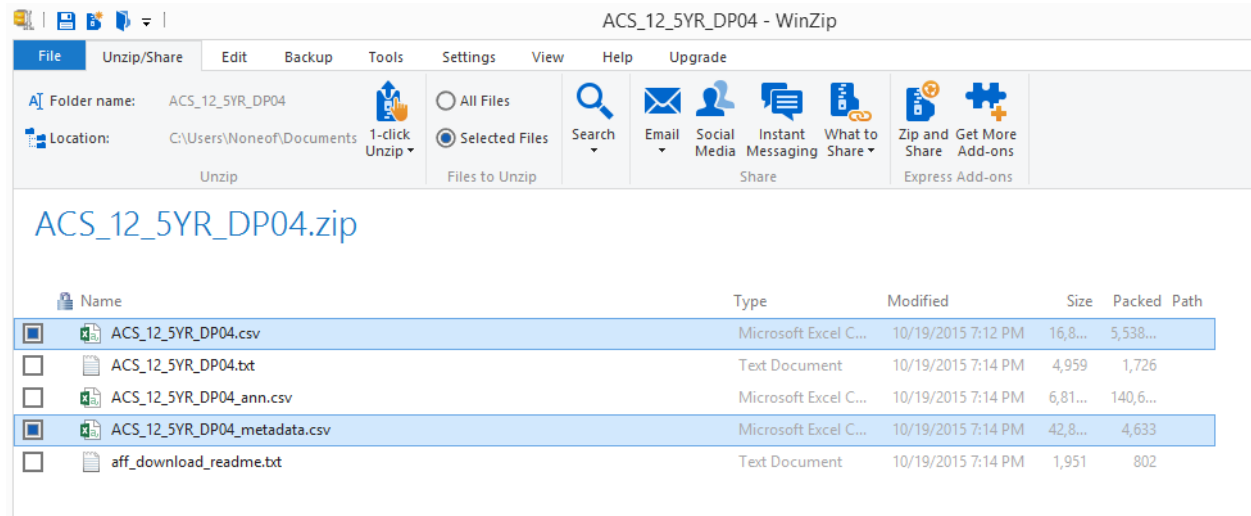
7. Click Download to save to your computer



8. Save the Download (zip file)



9. Open the download > click on data file (ACS\_12\_5YR\_TABLENUMBER.csv) and its data dictionary (ACS\_12\_5YR\_TABLENUMBER\_metadata.csv) > drag and drop files into your desired folder



**Appendix B. Download Instructions for Statewide Election Data**

1. For the 2010 General Election, go to [http://statewidedatabase.org/d10/g10\\_registration.html](http://statewidedatabase.org/d10/g10_registration.html)
2. Click on statewide Registration," REG block" to download .dbf file, on total counts (TOTREG) of persons registered. The .dbf file can be opened in Excel 2013
3. Click on statewide Registration," NOVOTE" to download .dbf file, on total counts (TOTREG) of persons registered, but did not vote. The .dbf file can be opened in Excel 2013

**2010 GENERAL ELECTION REGISTRATION DATA**

Below you will find 2010 census block data for the 2010 general registration.

Please click here for [precinct data file \(SOV, REG, ABS, POLLV, & VOTE\) content descriptions](#).

There are 58 files for each file type, one for each county. Files are named using county codes 01 to 58, please refer to the [county code list](#) to look up the code for a specific county.

Descriptions of the variable codes used in the data files can be found in the [technical documentation](#). The relevant appendix begins on page 27.

	REG	ABS	POLLVOTE	VOTE	NOVOTE
CALIFORNIA STATEWIDE	block	block	block	block	block

4. For 2010, downloaded files (statewide\_reg\_stats\_by\_block.dbf, statewide\_novote\_stats\_by\_block.dbf) were opened in Excel2013, and saved in .csv format )

5. For 2012, go to [http://statewidedatabase.org/d10/g12\\_geo\\_conv.html](http://statewidedatabase.org/d10/g12_geo_conv.html) to get the crosswalk of precincts to 2010 census blocks
6. Download the crosswalk RGPREC TO 2010 BLK zip file, and unpack "state\_g12\_rg\_blk\_map.dbf", open in Excel and save as a .csv file "state\_g12\_rg\_blk\_map.csv"

The screenshot shows a web browser window with the URL [http://statewidedatabase.org/d10/g12\\_geo\\_conv.html](http://statewidedatabase.org/d10/g12_geo_conv.html). The website header features the text "Statewide Database" and a navigation menu with "HOME", "ABOUT", and "SITEMAP".

**2012 GENERAL ELECTION GEOGRAPHIC DATA**

From this page you can download precinct boundary files and geographic conversion files. Descriptions of each of these files can be found by following the links below. Please note based on U.S. Census Bureau 2010 TIGER/Line census blocks.

**Precinct Boundary Files**

- [MPREC\\_CDF](#)
- [MPREC\\_MIF](#)
- [MPREC\\_SHP](#)
- [SRPREC\\_CDF](#)
- [SRPREC\\_MIF](#)
- [SRPREC\\_SHP](#)

**Data Conversion Files**

- [SRPREC TO 2010 BLK](#)
- [2010 BLK TO MPREC](#)
- [MPREC TO SRPREC](#)
- [SRPREC TO CITY](#)

Go to the [2012 General Election precinct data](#) for voter registration data and election results.

	MAP PRECINCTS	SR PRECINCTS	DATA CONVERSION
CALIFORNIA STATEWIDE	<ul style="list-style-type: none"> <li>by <a href="#">MPREC_CDF</a></li> <li>by <a href="#">MPREC_MIF</a></li> <li>by <a href="#">MPREC_SHP</a></li> </ul>	<ul style="list-style-type: none"> <li>by <a href="#">SRPREC_CDF</a></li> <li>by <a href="#">SRPREC_MIF</a></li> <li>by <a href="#">SRPREC_SHP</a></li> </ul>	<ul style="list-style-type: none"> <li><a href="#">SRPREC TO 2010 BLK</a></li> <li><a href="#">RGPREC TO 2010 BLK</a></li> <li>no data yet</li> <li>no data yet</li> <li>no data yet</li> <li><a href="#">SRPREC TO CITY</a></li> </ul>

7. To get the registered voters who voted, go to: <http://statewidedatabase.org/d10/g12.html>
8. Download the voter file with the REGPREC key

**2012 GENERAL ELECTION PRECINCT DATA**

From this page you can download Statement of Vote and Statement of Registration data by each type of file and precinct can be found by following the links below.

*Available Precinct data files (SOV, REG, ABS, MAIL, POLLV and VOTE.) & file content descriptions*

*Precinct type (rgprec, rrprec, srprec, ssprec, svprec, mprec etc.) descriptions*

*Go to the 2012 General Election geographic data, including the precinct boundary files.*

*\*The message "no data yet" indicates that the file is still being collected and processed. It v  
The message "unavailable" indicates that the data are not available from the county.*

*Dates of most recent files by county.*

	SOV	REG	ABS	MAIL	POLLV	VOTE
CALIFORNIA STATEWIDE	by svprec by srprec	codebook by rgprec by rrprec by srprec	codebook by rgprec by rrprec by srprec	codebook by rgprec by rrprec by srprec	codebook by rgprec by rrprec by srprec	codebook by rgprec by rrprec by srprec

6. Unpack "state\_g12\_voters\_by\_g12\_rgprec.dbf", open in Excel and save as a .csv file "state\_g12\_voters\_by\_g12\_rgprec.csv"

Note: the block map file has the total registered voters by block (REGBLK). To get the nonvoters, you must merge the block map to the voters file on the key REGPREC\_KEY. Multiply the percent of the precinct allocated to the block (PCTRGPREC) by the number of voters. See the R program for computational details.