UCDAVIS
CENTER FOR REGIONAL CHANGE

Regional Opportunity Index Overview

The Regional Opportunity Index (ROI) is comprised of two indices which assess the relative well-being of people and places for census tracts in the state of California. This document describes the data and the methods used to construct these indices, as well as data limitations that require caution when interpreting and using them.

The ROI is constructed using data from the American Community Survey (ACS), the Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) data, the California Department of Education, the California Department of Public Health, and several other data sources. To increase reliability of the estimates at the census tract level, we use the 5-year estimates from ACS, and combine multiple years of data for some of the other data sources. The actual years of data included in any given set of estimates is described on the ROI maps and in the downloadable data.

The People and Place indices encompass several domains, and each domain is composed of two or more indicators. This document provides a description of every indicator and the data sources and methods that were used to produce it. The indicators are combined using an averaging technique which calculates the geometric mean (described in the methodology section below) to create a domain score, and the domains are combined using the same technique to calculate the ROI: People and ROI: Place scores.

The indicators, domain scores, and index values are displayed on maps using color gradations to indicate census tracts experiencing more, or less, opportunity relative to other tracts in the region. Because of data limitations, we were not able to calculate the domains and indices in every census tract. Census tracts that are gray or are have an asterisk in them are marked in this way to indicate missing data or data that was considered unreliabile and therefore was not used in index calculation. In these tracts, caution must be used when interpreting the findings. Further information about data limitations below can be found at the end of this document.

A full version of the data, including the original indicators and calculated indices, is available for download in excel format from the Regional Opportunity Index website: http://interact.regionalchange.ucdavis.edu/roi/data.html.

Contents

Regional Opportunity Index: People 3
Education Opportunity: People 3
Economic Opportunity: People 5
Housing Opportunity: People 5
Mobility/Transportation Opportunity: People 6
Health/Environment Opportunity: People 7
Civic Life Opportunity: People 8
Regional Opportunity Index: Place 9
Education Opportunity: Place 9
Economic Opportunity: Place 11
Housing Opportunity: Place 13
Mobility/Transportation: Place 13
Health/Environment Opportunity: Place 14
Civic Life Opportunity: Place 15
Other Data 16
Methodology 22
Variable Selection 22
Availability and Currency 22
Geographic Scale 22
Reliability 23
Variable Transformations 25
z-Scores 25
Min-Max Scaling 25
Index Calculation 26
Geometric Mean 26
Domain Means and Index Values 27
How to Interpret Maps and Popup Bar Charts 27
Limitations 29
Weighting 29
Reliability 29

Validity 29

Regional Opportunity Index: People

The Regional Opportunity Index (ROI): People is a relative measure of people's assets in education, the economy, housing, mobility/transportation, health/environment, and civic life.

Education Opportunity: People

The Education Opportunity: People domain assesses people's relative success in gaining educational assets, in the form of a higher education, elementary school achievement, and regular elementary school attendance.

Education-People: College-Educated Adults (%)

Percentage of adults age 25 and over who have completed a post-secondary certificate/degree .

Calcuation: Number of adults age 25 or older who have completed an Associate's degree or higher divided by the number of adults age 25 or older, multiplied by 100. The category "Associate's degree" includes people whose highest degree is an associate's degree, which generally requires 2 years of college level work and is either in an occupational program that prepares them for a specific occupation, or an academic program primarily in the arts and sciences. The course work may or may not be transferable to a bachelor's degree. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B15002

Education-People: English Proficiency (%)

Three-year average of percentage of 4th graders who scored proficient or above on the English Language Arts portion of California's Standardized Testing and Reporting (STAR) test.

Calcuation: CDE reports the percentage of students who scored at the proficient or advanced levels on the English Language Arts section of the STAR test, out of those who took the test. We calculate the three-year average of this percentage for 4th graders in each school. If data is missing for any one year, we use the data from the remaining years to calculate the school's three-year average. For each tract, we take the average of the school means for the three elementary schools closest to the tract center. If any of the three closest schools are missing data, we use data for the remaining schools.

Only traditional elementary and K-12 schools (School Ownership Codes (SOC) 60, 61, and 65), including charter schools, are used in the calculation. Schools that closed prior to July 1 of the final year of the reference period are excluded from the calculation. Refer to the maps or the downloadable data to determine what years are included in the reference period for a specific version of the ROI.

We determined school proximity by calculating the distance from the population-weighted tract center to the geocoded location of each school using ArcGIS software. Note that this calculation is based on straight-line distance, and does not account for road placement or geographical features which may lengthen actual travel distance. Moreover, we did not account for school district boundaries, so it is possible that one or more schools included in the tract's average are located in districts other than the one that covers the tract center.

Source: California Department of Education STAR Test Results Research Files

Education-People: Math Proficiency (%)

Three-year average of percentage of 4th graders who scored proficient or above on the math portion of California's Standardized Testing and Reporting (STAR) test.

Calcuation: CDE reports the percentage of students who scored at the proficient or advanced levels on the math section of the STAR test, out of those who took the test. We calculate the three-year average of this percentage for 4th graders in each school. If data is missing for any one year, we use the data from the remaining years to calculate the school's three-year average. For each tract, we take the average of the school means for the three elementary schools closest to the tract center. If any of the three closest schools are missing data, we use data for the remaining schools.

Only traditional elementary and K-12 schools (School Ownership Codes (SOC) 60, 61, and 65), including charter schools, are used in the calculation. Schools that closed prior to July 1 of the final year of the reference period are excluded from the calculation. Refer to the maps or the downloadable data to determine what years are included in the reference period for a specific version of the ROI.

School proximity was determined by calculating the distance from the population-weighted tract center to the geocoded location of each school included the data using ArcGIS software. Note that this calculation is based on straight-line distance, and does not account for road placement or geographical features which may lengthen actual travel distance. Moreover, we did not account for school district boundaries, so it is possible that one or more schools included in the tract's average are located in districts other than the one that covers the tract center.

Source: California Department of Education STAR Test Results Research Files

Education-People: Elementary Truancy Rate (%)

Three-year average of percentage of students who have missed more than 30 minutes of instruction without an excuse at least three times during the school year.

Calcuation: The truancy rate is the number of students having unexcused absences of more than 30 minutes on three or more days, divided by the total number of students as reported on information day in October, multiplied by 100. The truancy rate is capped at 100% but may exceed that if cumulative truancy counts over the course of the school year exceeds total enrollment on information day. The calculated value is the mean of the three-year average annual truancy rate for the three elementary schools closest to the tract center. This indicator is inverted for the index by subtracting it from 100%.

Traditional elementary and K-12 schools (School Ownership Codes (SOC) 60, 61, and 65) that serve 4th grade students, including charter schools, are used in the calculation. Schools that closed prior to July 1 of the final school year of the reference period are excluded from the calculation. Refer to the maps or the downloadable data to determine what years are included in the reference period for a specific version of the ROI.

School proximity was determined by calculating the distance from the population-weighted tract center to the geocoded location of each school included the data using ArcGIS software. Note that this calculation is based on straight-line distance, and does not account for road placement or geographical features which may lengthen actual travel distance. Moreover, we did not account for school district boundaries, so it is possible that one or more schools included in the tract's average are located in districts other than the one that covers the tract center.

Source: California Department of Education, DataQuest Expulsion, Suspension, and Truancy report (2009 and 2010) and CBEDS Data About Schools/Districts (2011 and beyond)

Economic Opportunity: People

The Economic Opportunity: People domain measures the relative economic well-being of the people in a community, in the form of employment and income level.

Economy-People: Employment Rate (%)

Percentage of adults age 20-64 who are employed.

Calculation: The number of civilian adults age 20-64 who are employed divided by the umber of civilian adults age 20-64 in the labor force, multiplied by 100. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B23001

Economy-People: Minimum Basic Income (%)

Percentage of people with income over 200% of the federal poverty level.

Calcuation: Percentage of tract population (for which poverty status was determined) with income over 200% of the FPL divided by the population for whom poverty status was determined, multiplied by 100. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table C17002

Housing Opportunity: People

The Housing Opportunity: People domain measures the relative residential stability of a community, in the form of homeownership and housing costs.

Housing-People: Homeownership (%)

Percentage of housing units which are owned by their occupants.

Calcuation: Number of owner-occupied housing units divided by total occupied housing units, multiplied by 100.

A housing unit is a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied (or if vacant, is intended for occupancy) as separate living quarters. Separate living quarters are those in which the occupants live separately from any other people in the building and which have direct access from the outside of the building or through a common hall. The occupants may be a single family, one person living alone, two or more families living together, or any other group of related or unrelated people who share living arrangements. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B25003

Housing-People: Housing Cost Burden (%)

Percentage of households paying less than 30% of household income on housing costs.

Calcuation: Number of households whose housing costs are less than 30% of household income in the past 12 months divided by total occupied housing units, multiplied by 100. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B25106

Mobility/Transportation Opportunity: People

The Mobility/Transportation Opportunity: People domain contains indicators that assess a community's relative opportunities for overcoming rural isolation.

Mobiility/Transportation-People: Commute Time (%)

Percentage of workers whose commute time is less than 30 minutes.

Calculation: Number of workers age 16 and older whose commute time to work is less than 30 minutes, divided by the number of workers age 16 and older who don't work at home, multiplied by 100. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B08303

Mobility/Trans-People: Vehicle Availability (%)

Percentage of households with at least 1 vehicle, or 1 vehicle per worker.

Calcuation: Number of households with at least 1 vehicle per worker, or at least 1 vehicle if no workers, divided by total number of households, multiplied by 100. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B08203

Mobility/Trans-People: Internet Access (#)

Number of households per 1,000 with high-speed internet, categorized as 0 (none), 1 (1-199 households per 1,000 with high-speed internet), 2 (200-399) 3 (400-599) 4 (600-799) 5 (800 or more).

Calcuation: High-speed internet is defined by the FCC as service that is more than 200kpbs in at least one direction. The FCC categorizes tracts by the number of households per 1,000 that have high-speed internet (HSI). The categories are 0 (zero households with HSI), 1 (1-199 households per 1,000 with HSI),2 (200-399 households per 1,000 with HSI),3 (400-599 households per 1,000 with HSI), 4 (600-799 households per 1,000 with HSI), and 5 (800 or more households per 1,000 with HSI).

Source: FCC June 2013

Health/Environment Opportunity: People

The Health/Environment Opportunity: People domain measures the relative health outcomes of the people within a community, in the form of infant and teen health and general health.

Health/Environment-People: Infant Heath (%)

Three-year average of percentage of healthy birthweight babies.

Calcuation: The indicator is the total number of singleton births with a weight at or above 2500 grams (about 5.5 pounds) over the three year reference period, divided by the total number of singleton births during that same period, multiplied by 100. Birth records were geocoded to the census tract of the mother's residence. Approximately 4% of addresses could not be geocoded; these records were dropped. Refer to the maps or downloadable data for information about which years of data were used for a specific version of ROI. Values in tracts with fewer than 25 births in that time period are considered unreliable and should be interpreted with caution.

Source: California Department of Public Health, Birth Statistical Master Files

Health/Environment-People: Births to Teens (%)

Three year average of percentage of all births that were to teens.

Calcuation: Counting multiple births as one birth event , the indicator is the number of births to women under the age of 20 during the the three-year reference period, divided by the total number of births in that same time period, multiplied by 100. The indicator is inverted for the index by subtracting it from 100. Birth records were geocoded to the census tract of the mother's residence. Approximately 4% of addresses could not be geocoded; these records were dropped. Refer to the maps or downloadable data for information about which years of data were used for a specific version of ROI. Values in tracts with fewer than 25 births in that time period are considered unreliable and should be interpreted with caution.

Source: California Department of Public Health Birth Statistical Master Files

Health/Environment-People: Years of Life Lost Rate

Three year average of the years of potential life lost rate per 1,000 population under the age of 65.

Calcuation: Years of potential life lost (YPLL) is a measure of premature death, or the number of years of life lost among those who died before a predetermined age. We set that age at 65 to assess the number of prime working years lost, assuming an average retirement age of 65. YPLL is calculated by subtracting the age at death from 65 for all deaths that occurred before the age of 65 and ignoring those at or above age 65, summing the results. Death records were geocoded to the census tract using the decedant's residential address. Less than 4% of addresses could not be geocoded; these records were dropped.

For this indicator, we calculate YPLL over fixed age categories, using the mid-point of the category's age range to determine YPLL for that category, then multiplying by the number of deaths in that age range. For example, for deaths to persons between ages 40 and 45, the midpoint is 42.5 and YPLL for all persons in this group is 65 – 42.5 = 22.5. We multiple the YPLL of 22.5 by the annual average number of deaths in this age group over the three-year reference period to get the total YPLL for this age group. We repeat this step for all age groups, summing the YPLL over all age groups to arrive at the total YPLL. The YPLL rate is this total divided by the population under the age of 65. The result is then divided by 1,000 to arrive at the YPLL rate. Refer to the maps or downloadable data for information about which years of data were used for a specific version of ROI.

Source: California Department of Public Health Death Statistical Master Files; Census 2010, SF1 Table P12

Civic Life Opportunity: People

The Civic Life Opportunity: People domain measures the relative social and political engagement of an area, in the form of households that speak English and voter turnout.

Civic Life-People: Voting Rates (%)

Percentage of citizen voting age population that voted in a statewide General Election.

Calculation: Number who voted in a California General Election divided by the number of citizens of voting age (CVAP) multiplied by 100. Refer to the maps or downloadable data to determine which years of data are used for a specific version of the ROI.

CVAP estimates were obtained from a special tabulation of the ACS, which can be downloaded from http://www.census.gov/rdo/data/voting_age_population_by_citizenship_and_race_cvap.html.

Source: California Registrar of Voters, General Election Statement of Registration; ACS 5-year estimates, CVAP Special Tabulation

Civic Life-People: English Speakers (%)

Percentage of population age 18-64 that speaks only English or speaks English "well" or "very well"

Calculation: Sum of the number age 18-64 who speak only English, who speak English "well" and those who speak English "very well", divided by the total number age 18-64, multiplied by 100. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B16004

Regional Opportunity Index: Place

The Regional Opportunity Index (ROI): Place is a relative measure of an area's assets in education, the economy, housing, mobility/transportation, health/environment, and civic life.

Education Opportunity: Place

The Education Opportunity: Place domain assesses a census tract's relative ability to provide educational opportunity, in the form of high-quality schools that meet the basic educational and social needs of the population.

Education-Place: High School Graduation Rate (%)

Three-year average of percentage of 9th grade cohort that graduated from high school four years later.

Calculation: The graduation rate indicator is the the number of students in a 9th grade cohort that graduated from high school within four years, divided by the size of the cohort, multiplied by 100, and averaged over three years. The graduation rate is based on district level data for Unified and High School districts that have District Ownership Codes (DOC) of 52 or 54. These reports include all schools in the districts, including alternative schools of choice, continuation high schools and community day schools. Refer to the maps or the downloadable data to determine what years are included in the reference period for a specific version of the ROI.

Districts are mapped to census tracts using the MABLE/Geocorr12: Geographic Correspondnce Engine. If a tract is covered by more than one district, thevalue assigned to the tract is the weighted average of each district's graduation rate, where the weights are determined by the portion of the tract population covered by each district.

Source: California Department of Education, Cohort Outcome Data; Missouri Census Data Center, MABLE/Geocorr12 Version 1.2

Education-Place: UC/CSU Eligibility (%)

Three-year average of the percentage of high school graduates who completed UC/CSU a-g course requirements.

Calculation: The college ready rate is the number of high school graduates in a school district who completed the UC/CSU a-g course requirements divided by the total number of high school graduates in the district, multiplied by 100, and averaged over three years. Annual district means are calculated by combining graduation data from all schools in Unified and High School

districts that have School Ownership Codes (SOC) between 60 and 69 inclusive. In addition to traditional schools, this includes alternative schools of choice, continuation high schools and community day schools. Schools that closed prior to July 1 of the final school year included in the reference period are excluded from the calculation. Refer to the maps or the downloadable data to determine what years are included in the reference period for a specific version of the ROI.

Districts are mapped to census tracts using the MABLE/Geocorr12: Geographic Correspondnce Engine. If a tract is covered by more than one district, the value assigned to the tract is the weighted average of each district's graduation rate, where the weights are determined by the portion of the tract population covered by each district.

Source: California Department of Education, Graduates by Race & Gender; Missouri Census Data Center, MABLE/Geocorr12 Version 1.2

Education-Place: Teacher Experience (%)

Three-year average of percentage of teachers at the three closest public elementary schools with more than 5 years of teaching experience and at least one year of education beyond a BA.

Calculation: The number of teachers in a school who have more than 5 years of teaching experience and have at least one year of education beyond a BA, divided by the total number of teachers in then school, multiplied by 100. The indicator is the mean of the three-year average annual percentage of experienced teachers for the three elementary schools closest to the tract center. If any of the three closest schools are missing data, data for the remaining schools are used. Based on data for traditional elementary and K-12 schools, including charter schools (School Ownership Codes (SOC) 60, 61, and 65). Schools that closed prior to July 1 of the final school year included in the reference period are excluded from the calculation. Refer to the maps or the downloadable data to determine what years are included in the reference period for a specific version of the ROI.

School proximity was determined by calculating the distance from the population-weighted tract center to the geocoded location of each school included the data using ArcGIS software. Note that this calculation is based on straight-line distance, and does not account for road placement or geographical features which may lengthen actual travel distance. Moreover, we did not account for school district boundaries, so it is possible that one or more schools included in the tract's average are located in districts other than the one that covers the tract center.

Source: California Department of Education, Staff Demographics

Education-Place: High School Discipline Rate (%)

Average annual percentage of high school students in the school district who were suspended or expelled.

Calculation: The total number of students in a district who were suspened or expelled, divided by total district enrollment, multiplied by 100, and averaged over three years. District suspension/expulsion rates are calculated by combining discipline data from all schools in the districts that have School Ownership Codes (SOC) between 60 and 69 inclusive. In addition to traditional schools, this includes alternative schools of choice, continuation high schools and community day schools. Schools that closed prior to July 1 of the final school year included in the reference period are excluded from the calculation. Refer to the maps or the downloadable data to determine what years are included in the reference period for a specific version of the ROI.

Students are counted more than once if they were suspended or expelled multiple times for different incidents. Thus, the number of suspensions and expulsions can exceed total enrollment. In those cases, the value is capped at 100%. This indicator is inverted for the index by subtracting it from 100%.

Districts are mapped to census tracts using the MABLE/Geocorr12: Geographic Correspondnce Engine. If a tract is covered by more than one district, the value assigned to the tract is the weighted average of each district's graduation rate, where the weights are determined by the portion of the tract population covered by each district.

Note: A reporting change in 2011 limits our ability to combine data that spans that year, meaning that we can use only two years of data, rather than three, for some versions of the ROI. Prior to 2011, school districts reported the total number of offenses committed by each student, meaning that a student could be counted more than once for the same expulsion or suspension. Starting in 2011, only the most serious offense committed by a student was reported. Thus it is not advisable to combine data that spans the reporting change. Refer to the maps or the downloadable data to determine what years are included in the reference period for a specific version of the ROI.

Source: California Department of Education, DataQuest Expulsion, Suspension, and Truancy report (2009 and 2010) and Expulsion and Suspension Data (2011 and beyond); Missouri Census Data Center, MABLE/Geocorr12 Version 1.2

Economic Opportunity: Place

The Economic Opportunity: Place domain measures the relative economic climate of a community, in the form of access to employment and business climate.

Economy-Place: Job Availability (#)

Number of jobs per 1,000 people, within a 5-mile radius.

Calculation: Number of workers per census block was extracted from the LEHD (Longitudinal Employer-Household Dynamics) Origin-Destination Employment Statistics (LODES) database for a single year. The number of workers employed in census blocks that lie within a 5 mile radius of

the tract center was divided by the census population aged 20-64 within those same census blocks divided by 1,000. Refer to the maps or downloadable data for information about which data year was used for a specific version of ROI. For more information about the aggregation of data over an area encircling the tract center, read about Geographic Scale in the Methodology section. Values in tracts with fewer than 100 jobs within the 5-mile radius of the tract center are considered unreliable and should be interpreted with caution.

Source: LODES; Census 2010, SF1 Table P12

Economy-Place: Job Quality (%)

Percentage of jobs that are in high-paying industries, within a 5-mile radius.

Calculation: Number of workers with high earnings (over $3,333 per month) was extracted from the LEHD (Longitudinal Employer-Household Dynamics) Origin-Destination Employment Statistics (LODES). The number of employees in these high paying jobs within a 5 mile radius of the tract center was divided by the total number of employees within the 5 mile radius and multiplied by 100. Refer to the maps or downloadable data for information about which data year was used for a specific version of ROI. For more information about the aggregation of data over an area encircling the tract center, read about Geographic Scale in the Methodology section. Values in tracts with fewer than 100 jobs within the 5-mile radius of the tract center are considered unreliable and should be interpreted with caution.

Source: LODES

Economy-Place: Job Growth (%)

Percentage 1-year change in the number of jobs, within a 5-mile radius.

Calculation: Business location and number of employees for two consecutive years was extracted from the LEHD (Longitudinal Employer-Household Dynamics) Origin-Destination Employment Statistics (LODES) database. To calculate a 1-year percentage change in number of jobs, the number of employees within a 5 mile radius of the tract center in the base year was subtracted from the number of employees within the same 5 mile radius the following year, divided by the total number of employees during the base year within the 5 mile radius, and multiplied by 100. Refer to the maps or downloadable data for information about which data years were used for a specific version of ROI. For more information about the aggregation of data over an area encircling the tract center, read about Geographic Scale in the Methodology section. Values in tracts with fewer than 100 jobs within the 5-mile radius of the tract center are considered unreliable and should be interpreted with caution.

Source: LODES

Economy-Place: Bank Accessibility (#)

Number of banks and credit unions per 1000 people, within a 5-mile radius.

Calculation: Number of banks (insured by the Federal Deposit Insurance Corporation [FDIC]) and credit unions (listed with the National Credit Union Association [NCUA]) within a 5-mile radius from the tract center divided by the total population within 5-miles, divided by 1000. Refer to the maps or downloadable data for information about which data year was used for a specific version of ROI. For more information about the aggregation of data over an area encircling the tract center, read about Geographic Scale in the Methodology section. For institutions insured by the FDIC, only "Full Service" banks have been included in the analysis.

Source: FDIC; NCUA; Census 2010, SF1 Table P12

Housing Opportunity: Place

The Housing Opportunity: Place domain measures relative availability of housing in a community, in the form of housing sufficiency and housing affordability.

Housing-Place: Housing Adequacy (%)

Percentage of households with no more than one occupant per room.

Calculation: Number of households with 1 or fewer occupants per room divided by total occupied housing units, multiplied by 100. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B25014

Housing-Place: Housing Affordability

Ratio of median income of census tract to median value of dwellings in census tract.

Median household income divied by median value of owner occupied dwellings. ACS caps home prices at $1,000,001, but not provide a margin of error (MOE) for tracts where this was done. We calculated the average MOE for tracts with median home prices ranging from $900-999,000, and assigned that MOE to tracts with median home values over $1 million. We cap the housing affordability ratio at 0.50 to reduce the effect that 7 outliers (tracts with extremely high values compared to the rest of the tracts) have on the index. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Tables B19013 and B25077

Mobility/Transportation: Place

No indicators have been identified at this time.

Health/Environment Opportunity: Place

The Health/Environment Opportunity: Place domain is a relative measure of how well communities meet the health needs of their constituents, in the form of access to health care and other health-related environments.

Health/Environment-Place: Prenatal Care (%)

Three year average of percentage of mothers who received prenatal care in first trimester.

Calculation: The number of mothers who started receiving prenatal care in the first trimester of their pregnancy during the reference period, divided by the total number of births during the same period, multiplied by 100. Birth records were geocoded to the census tract of the mother's residence. Approximately 4% of addresses could not be geocoded; these records were dropped. Multiple births are included, though only counted as one birth event. Refer to the maps or downloadable data for information about which years of data were used for a specific version of ROI. Values in tracts with fewer than 25 births are considered unreliable and should be interpreted with caution.

Source: California Department of Public Health Birth Statistical Master Files

Health/Environment-Place: Distance to Supermarket (%)

Percentage who live within 1 mile (urban) or 10 miles (rural) of supermarket.

Calculation: USDA defines low access to a healthy food source as living more than 1 mile in urban areas and more than 10 miles in rural areas from a supermarket or large grocery store. USDA calculates the proportion of population with low access by dividing the geography into ½-km square grids, allocating 2010 Census population data to each square, and measuring distance between the geographic center of a populated square and the center of the square with the nearest supermarket. The ROI multiples the inverse of the proportion with low access by 100 to derive the percentage with access to a healthy food source. USDA defines supermarkets as stores that sell all major categories of food and have annual sales of at least $2 million, and uses 1) a list of stores authorized to receive Supplemental Nutrition Assistance Program (SNAP) benefits, and 2) a proprietary supermarket store listing from Trade Dimensions TDLinx (a Nielsen company). The USDA urban versus rural designation varies from US Census in that, rather than a designation based on the geographic centroid of a tract, the USDA uses the population-weighted centroid.

Source: USDA Food Access Research Atlas 2010

Health/Environment-Place: Health Care Availability (#)

Number of providers of basic medical services per 1000 population within 5 mile radius.

Calculation: The location of establishments that provide basic medical health care services – those with North American Industry Classification System (NAICS) codes of Offices of Physicians (6211) or General Medical and Surgical Hospitals (6221) - was extracted from the National

Establishment Time-Series (NETS) database for a single year. The indicator divides the number of health care establishments within a 5 mile radius in that year from the population-weighted center of a census tract by the total population within a 5-mile radius, divided by 1000. Refer to the maps or downloadable data for information about which NETS data year was used for a specific version of ROI. For more information about the aggregation of data over an area encircling the tract center, read about Geographic Scale in the Methodology section.

Source: NETS; Census 2010, SF1 Table P12

Health/Environment-Place: Air Quality (PM2.5)

Annual mean concentration of PM2.5.

Calculation: Annual mean concentration of particles that are less than 2.5 micrometers in diameter. OEHHA computes the variable by taking the mean of quarterly means for monitors within 50km of the tract centroid. Monitors which reported fewer than 75% of the expected number of observations, based on scheduled sampling frequency, were dropped from the calculation. Tracts that were further than 50km from the nearest monitor are missing data.

The EPA's standard for annual PM2.5 concentration is 15 micrograms per cubic meter (µg/m3). Exposure to concentrations above this level can aggravate heart and lung disease, resulting in increased hospital admissions, emergency room visits, absences from school or work, restricted activity days, and premature death.

Source: Office of Environmental Health Hazard Assessment (OEHHA) and Air Resources Board (CalEPA)

Civic Life Opportunity: Place

The Civic Life Opportunity: Place domain measures the relative social and political stability of an area, in the form of neighborhood stability (living in same residence for one year) and US citizenship.

Civic Life-Place: Neighborhood Stability (%)

Percentage of citizens, over age 1, who live in the same residence as the previous year.

Calculation: Number of people in the same house as the previous year divided by population over age 1, multiplied by 100. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B07001

Civic Life-Place: US Citizenship (%)

Percentage of adults who are U.S. citizens.

Calculation: Number over 18 who are citizens (native and naturalized) divided by population over age 18, multiplied by 100. Refer to the maps or downloadable data to determine which ACS release is used for a specific version of the ROI.

Source: ACS 5-year estimates, Table B05003

Other Data

In addition to the ROI indicators and indices, we provide other data that may be useful to parties who want to know more about the distribution of opportunity across the region.

Other: Demographic

Demographic information is available by census block group and census tract. The block group data are obtained from the 2010 Census, while the tract data are estimates from the ACS. We map total population, percent of population under age 18, percent of population over 64, percent of population that is Hispanic, percent of population that is non-Hispanic Asian or Pacific Islander, percent of population that is non-Hispanic African American, percent of population that is non-Hispanic White, and percent that is all other (this includes American Indian/Alaskan Native, Other, and two or more categories).

Source: Census 2010, SF1 Table T11 and Table P5; ACS 5-year estimates, Tables B01001 and B03002

Other: Rural-Urban Commuting Area (RUCA) Code

A four-category characterization of the primary commuting destination for tract residents: 1) metropolitan area, 2) micropolitan area, 3) small town, 4) rural area.

The USDA's Economic Research Service Rural-Urban Commuting Area (RUCA) codes identify urban cores and adjacent territory that is economically integrated with those cores. The most recent RUCA codes are based on population density, urbanization, and daily commuting data from the 2010 decennial census and the 2006-10 American Community Survey. The Census Bureau identifies two types of urban areas based on population density:

Urbanized Areas (UAs) of 50,000 or more people;

Urban Clusters (UCs) of at least 2,500 and less than 50,000 people.

"Rural" encompasses all population, housing, and territory not included within an urban area. RUCA codes delineate metropolitan, micropolitan, small town, and rural commuting areas based on the size and direction of the primary (largest) commuting flows.

The ten primary RUCA codes are defined as follows:

1) Metropolitan area core: primary flow within an urbanized area (UA);

2) Metropolitan area high commuting: primary flow 30% or more to a UA;

3) Metropolitan area low commuting: primary flow 10% to 30% to a UA;

4) Micropolitan area core: primary flow within an Urban Cluster of 10,000 to 49,999 (large UC);

5) Micropolitan high commuting: primary flow 30% or more to a large UC;

6) Micropolitan low commuting: primary flow 10% to 30% to a large UC;

7) Small town core: primary flow within an Urban Cluster of 2,500 to 9,999 (small UC);

8) Small town high commuting: primary flow 30% or more to a small UC;

9) Small town low commuting: primary flow 10% to 30% to a small UC;

10) Rural areas: primary flow to a tract outside a UA or UC.

We collapse these 10 codes into 4 based on the primary commuting flow: codes 1-3, codes 4-6, codes 7-9, and code 10 are recoded as codes 1 (metropolitan), 2 (micropolitan), 3 (small town), and 4 (rural) respectively.

Tracts with zero population are not assigned a RUCA code. The ERS also produces a set of secondary codes based on secondary commuting flows but we do not use those here.

Source: USDA Economic Research Service 2013

Other: Economic Dependence Code

The USDA has classified all U.S. counties according to six non-overlapping categories of economic dependence based on earnings by place of work. The categories are 1) farming, 2) mining, 3) manufacturing, 4) services, 5) Federal/State government, and 6) unspecialized counties.

The USDA describes their methodology in this document, a portion of which is reproduced below.

Methods for Determining the Economic Dependence Types

Labor and proprietors' earnings by place of work are the basis for the economic dependence categories. Each industry's earnings were calculated as a percent of total labor and proprietors' earnings in the county in 1998, 1999, and 2000. These percentages were summed, and divided by 3 to obtain annual average percentages. This averaging was done to minimize the effects of any one-year anomaly in an industry's earnings. For simplicity, all labor and proprietors' earnings in a county are referred to as total county earnings.

County-level estimates of earnings by place of work used to measure economic dependence came from the Bureau of Economic Analysis' (BEA) Regional Economic Information System (REIS). BEA recalculated state and county earnings for all years in its REIS when it released new 2002 data in May 2004. The years 1969-2001 were revised from the previous release of May 2003. These new estimates incorporate the results of

comprehensive revision to the national income and product accounts released December 10, 2003 and to state personal income released April 27, 2004. The revised estimates also reflect new and revised county-level source data.

Selection of the industries ERS classified was guided by regional economics theory. Farming, mining, manufacturing, and Federal/State government industries produce goods or services for export outside the local economy. Exporting industries are termed 'basic' in regional economics and are often shown to be sources of larger growth in local economies (or declines during economic downturns) than industries that produce for the local market. Service industries may either produce for the local or export economies. ERS set a high service earnings threshold to help assure that the counties we classified as services-dependent do have service industries that serve more than the local population. These economic dependence categories are mutually exclusive.

Farming dependence was based on two thresholds-farm earnings accounting for an annual average of 15 percent or more of total county earnings during 1998-2000 or farm occupations accounting for 15 percent or more of all occupations of employed county residents in 2000. The farming occupation option was adopted to allow counties into the farming-dependent group that had highly farming-oriented economies but did not meet the earnings threshold, most often due to negative farm earnings estimates for some or all of the analyzed years. Farming dependence was determined first and takes precedence over all the other economic dependence types.

The final farming-dependent counties differ from the preliminary ones we published in May 2004, based on older BEA data. Nationally, BEA's revised county earnings estimates for farming are 12, 7, and 9 percent higher in 1998, 1999, and 2000 than in the older data release. By State, the revised estimates for those years also differ from the older data, but are not always higher. For example, in 2000, Minnesota's revised farm earnings are 19 percent less than the old estimate while Wisconsin's revised farm earnings are more than twice the old estimate. Even with such large national and State revisions, only 35 counties differ in their final farming-dependent status from their preliminary status. Twenty-seven lost their preliminary farming-dependent status and 8 counties gained final farming-dependent status. Of the preliminary farming-dependent counties, 432 (94 percent) remain farming-dependent in the final codes.

Mining (including metal; coal; oil and gas; stone; sand and gravel; clay, ceramic, and refractory minerals; chemical and fertilizer minerals; and miscellaneous nonmetallic minerals, such as gem stones, diatomaceous earth, peat, and talc) and **Federal/State government** dependence were also based on the industry accounting for an annual average of 15 percent or more of total county earnings during 1998-2000.

Manufacturing dependence was based on accounting for an annual average of 25 percent or more of total earnings during the 3 years.

Services dependence (including retail trade, finance, insurance, real estate, and services as defined by the Standard Industrial Classification System (SIC)) was based on accounting for an annual average of 45 percent or more of total earnings during the 3 years.

If a county qualified for more than one of mining, Federal/State government, or manufacturing types, it was classified in the industry in which it was the largest number of percentage points above the threshold. Services were not allowed to take such precedence over the other three industries. There were a few counties in which services exceeded its 45-percent threshold by more than other industries exceeded their thresholds. Most of those counties were State university or capital counties where we believe the service industries follow from the Federal/State government industry being concentrated there rather than government following services.

Counties that are not classified as dependent upon any of those industries are termed **Nonspecialized**. ERS has not explored whether any other particular industry is concentrated in those counties. BEA began reporting industry data by the North American Industry Classification System (NAICS) with its 2001 REIS data. In a few years, we plan to revisit the economic dependence classification, using NAICS industries to do a more detailed analysis of industry concentrations. That analysis will probably leave many fewer nonspecialized counties.

Source: USDA Economic Research Service, County Typology Codes 2015

Other: City Revenues from Local Sources (%)

Percentage of city revenues from local versus federal or state sources

Calculation: Using the CA Controller's Office Cities Financial Transactions Report, the amount of total revenue from local and county sources is divided by total revenue from all sources and multiplied by 100 to derive the percentage of city revenues from local sources. Because the data pertains to incorporated cities, not census tracts, fiscal autonomy is excluded from the ROI calculation, but is presented as a supplemental layer. Refer to the maps or downloadable data to determine what data year is used for a specific version of the ROI.

Source: California State Controller's Office, Division of Accounting and Reporting, Cities Financial Transactions Report

Other: City revenues from local sources per capita ($)

City revenues from local sources per capita.

Calculation: Using the CA Controller's Office Cities Financial Transactions Report, the amount of total revenue from local and county sources is subtracted from total revenue from all sources and divided by the total population. Because the data pertains to incorporated cities, not census tracts, fiscal autonomy is excluded from the ROI calculation, but is presented as a supplemental layer. Refer to the maps or downloadable data to determine what data year is used for a specific version of the ROI.

Source: California State Controller's Office, Division of Accounting and Reporting, Cities Financial Transactions Report; ACS 5-year estimates, Table B01003

Other: Tracts with Low Populations and/or High Prison Populations

Tracts that have very low populations or in which a large percentage of residents live in a prison are flagged with a value of 1 and displayed in gray in the map. **Calculation:** Tracts with total population below 100, fewer than 100 households, and those that contain a correctional facility in which more than 25% of residents live in correctional facilities are flagged with a value of 1.

Source: ACS 5-year estimates, Tables B01001 and B25002; ACS Prison Policy Initiative Correctional Facility Locator 2010 (http://www.prisonersofthecensus.org/locator2010/), 2010 Census SF1, Table P42

Other: Jobs-Housing Fit, Jurisdiction Level

The Jobs Housing Fit jurisdiction level ratio is the ratio of low-wage jobs (paying under $1250/month) to affordable rental units (less than $750/month rent) within the boundaries of cities and census designated places.

Calculation: Housing data is from the American Community Survey 5-year estimates is matched with jobs data from the Longitudinal Employer Household Dynamics (LEHD) Origin-Destination Employment Statistics Dataset (LODES), Workplace Area Characteristics file, published by the U.S. Census and available for download here: http://lehd.ces.census.gov/data/. It includes all employment covered by the Unemployment Insurance system, along with Federal Government employment. It excludes self-employed workers. Since its reference point is essentially jobs held on April 1st each year, it undercounts seasonable employment in other times of the year, which is especially relevant for the San Joaquin Valley, which has high levels of seasonal farm work that is not well captured in this dataset. We use the Workplace Area Characteristics file for the terminal year of the ACS 5-year estimates. For example, when paired ACS 2008-12 data with the LEHD 2012 data. Refer to the maps or downloadable data for the data years used in a particular version of the ROI.

The definition for low-wage jobs of $1250/month or less of earnings is pre-determined by the LODES dataset, which only reports on job earnings in three categories: earnings $1250/month or less; earnings $1251/month to $3333/month; and earnings greater than $3333/month.

In determining housing affordability, it was important for us to develop a threshold that was based on a multiple of this $1250 income threshold, rather than a measure of area median income (which is often used in affordable housing programs). This was because we want to be able to easily update the analysis on an annual basis and compare trends over time, and thus need a consistent measure of housing affordability that corresponds with the (unchanging) measure of low-wage jobs.

$750/month corresponds to the equivalent of 30% of household income if 2 income earners in a household were both earning $1250/month.

($750 * 2 * 30% = $750). This is probably a generous estimate of affordability, since the average household in California has approximately 1.4 income earners.

The thresholds of $150,000 for an affordable owned home is based on a calculation of monthly principal and interest payments on a 30-year 4% fixed-rate mortgage of $120,000 (80% of home-value) plus an estimated 1.2% general property tax and municipal assessments rate, which comes to $723/month.

This assumption doesn't take into account additional insurance costs or potential tax savings, and doesn't address where a 20% down-payment for the home might come from. Given these limitations in an assumption of owned-home affordability, our focus is on affordable rental units.

It is important to note that 'affordable housing' in this context does not refer to subsidized or deed-restricted units, which is frequently the definition used in the affordable housing field. Rather it is a measure of actual rent based on all units, regardless of deed restrictions or eligibility for subsidy. Developing the right fit between available housing types and the income level of households is an important part of regional planning and development. An imbalance in low-wage jobs and affordable housing is of concern not only for those low-wage workers who face challenges in finding affordable housing near work, but is of concern for regions as a whole, since it makes it more difficult to reduce overall vehicle miles travelled and potentially contributes to an excess fiscal burden on those jurisdictions with higher proportions of affordable apartments and houses.

Source: LODES; ACS 5-year estimates, Table B25056, Table B25061, Table B25075, Table B25085

Other: Jobs-Housing Fit, Tract Level

The Jobs Housing Fit tract level ratio is the ratio of low-wage jobs (paying under $1250/month) to affordable rental units (less than $750/month rent) within a 2.5 mile radius of the tract center.

Calculation: Like the Jurisdiction Level Jobs-Housing Fit measure, the Tract Level version of the measure is the ratio of low-wage jobs to affordable housing units, but it counts jobs and housing units that are within a 2.5 mile radius of the population-weighted center of the tract.

The population-weighted center can be thought of as the “average” location of all individuals in the tract. To determine the geographically-weighted centroid of each census tract, we used US 2010 Census population totals by census block, assigning the population count to the geographic center of the block, and then use this value to develop a population weighted centroid of the tract (using ArcGIS software to calculate the centroid). We then calculate a 2.5 mile radius around each tract’s population-weighted centroid, finding all census blocks which fall within that radius for each tract. The number of jobs and affordable housing units are aggregated over those census blocks to arrive at the total number within the 2.5 mile radius of the tract center.

Estimates of rents in the American Community Survey (ACS) are only available at a tract level. To estimate the number of affordable housing units within a 2.5 mile radius of the tract center, we needed information for geographic areas smaller than a census tract, which can be much larger

than 2.5 miles across in rural areas. The 2010 Census has data on housing units by census block. We used this data to calculate the proportion of a tract's housing units in each census block within the tract, and assume that the proportion of affordable units in each block is the same as the proportion of all units. This obviously may not be accurate in tracts that include diverse housing stock, but is a reasonable assumption for the analysis. We apply these proportions to the estimates of affordable housing units in the ACS data to arrive at the aggregate count of affordable housing units within a 2.5 mile radius of the tract center.

Source: LODES; ACS 5-year estimates, Table B25056, Table B25061, Table B25075, Table B25085

Other: Affordable Housing Developments

CalEnviroScreen 2.0

Methodology

Variable Selection

A review of the literature on community development and well-being identified certain key components of opportunity. The ROI domains correspond to these components. Within each domain, we looked for measures that would allow us to assess performance in that domain, and that met the following criteria.

Availability and Currency

The first criterion is that the data be readily available and updated regularly, so that we can track change in the ROI over time. For instance, we chose to use ACS data rather than decennial Census data, as it is updated annually, even though it is slightly less reliable due to smaller sample sizes (approximately one of every 8 persons is sampled for ACS 5-year estimates, compared to 1 in 6 for Census long-form questions). Almost all of the indicators we selected are updated annually, if not more frequently.

To determine what years of data are used in a particular version of the ROI, please refer to the maps or downloadable data.

Geographic Scale

Our second criterion is that data should permit evaluation at the census tract level, allowing users to assess opportunity at a relatively small geographic scale. Most of the indicators were based on data that was available at the tract level, with a few exceptions.

Recognizing that much activity occurs outside one's residential tract, some of our indicators are based on variables aggregated over a given radius from the tract center. By tract center, we mean the population weighted centroid, which can be thought of as the "average" location of all

individuals in the tract. To determine the tract centroid, we used US 2010 Census population totals by census block, assigning the population count to the geographic center of the block, and use this value to find the population weighted centroid of the tract (calculated using ArcGIS software). We then identify all census blocks that fall within a given radius for each tract, and aggregate the variable of interest over the census blocks so identified. For example, several of the Economic: Place indicators are based on the location of employers and the number of people they employ. We geocoded all employer locations to the census block level, then for each tract, summed the number of employers and employees in census blocks within a 5-mile radius of the population weighted tract centroid. Both numerators and denominators for the indicators based on this data are aggregated in this manner. Details about the specific variables used in the construction of such indicators can be found above.

Data from the California Department of Education is school-based, and while we know the location of schools, we don't know the residential location of students who attend those schools. In order to allocate school-level data to census tracts, we identified the three schools closest to the population-weighted center of each tract, and calculated the mean of the indicator for those three schools, and assigned the mean to the tract. School proximity was determined by calculating the distance from the population-weighted tract center to the geocoded location of each school using ArcGIS software. Note that this calculation is based on straight-line distance, and does not account for road placement or geographical features which may lengthen actual travel distance. Moreover, we did not account for school district boundaries, so it is possible that one or more schools included in the tract's average are located in districts other than the one that covers the tract center.

We used a slightly different tactic for the high school based indicators. Because student mobility is much higher at the high school level, we felt it would be inappropriate to use the same method of calculating tract-level indicators from school-level data. Instead, we calculate district averages for all high schools in each district, and apply the district average to all tracts that reside within the district. Almost all tracts reside completely within a single district; in the few cases where this is not true, we use a weighted average of the district means for the districts that cover the tract. The weights are derived from a district-tract crosswalk obtained from the University of Missouri's MABLE/Geocorr12 Version 1.2 and are simply the percent of the tract's population covered by each school district in the tract.

Reliability

Our second criterion is that the data be reasonably reliable, meaning that we have a relatively high level of confidence that the indicators are accurate representations of what is being measured. Reliability can be an issue when tracking events in small populations or in surveys based on small sample sizes. In these cases, estimates my not be reliable due to sample error or random fluctuations across time or space. We took steps to increase reliability where possible by combining multiple years of data (i.e., indicators based on Vital Statistics, CDE, and Cal/EPA PM2.5 data) or averaging over larger geographic areas to approximate tract-level processes (indicators based on CDE high school data, and NETS data).

For the same reason, we use ACS-5 year estimates rather than 1- or 3-year estimates. This effectively increases the sample size by aggregating all responses over the 5-year period, increasing the sampling fraction from about 1/40 to about 1/8. The resulting estimates have smaller standard errors, though it is important to remember that the longer time frame tends to smooth out short-term fluctuations and may mask rapidly occurring changes.

We evaluated all indicators to identify tracts in which the estimates do not meet an acceptable level of reliability. The threshold depends on how the indicator is measured, but we strive to assess reliability in terms of standard error of the estimate. Standard error is a measure of how much variation there is in the data, and is a function of sample size. The less variation in the data, and the larger the sample size, the more confidence we have that the estimate closely approximates the true value in the population. The more variation and the smaller the sample, the less confidence we have that the estimate is accurate.

One commonly used measure of reliability is Relative Standard Error (RSE), which is calculated by dividing the standard error by the estimate and then multiplying by 100. The RSE expresses the standard error as a percentage of the estimate itself. When the standard error is small relative to the estimate, the RSE will be small and our confidence in the estimate is high; when the standard error is large relative to the estimate, the RSE will be large and our confidence in the estimate low.

Data that are collected over an entire population (e.g. en entire high school cohort, all births, all deaths) are not subject to sampling error, but "the number of events that actually occurred may be considered as one of a large series of possible events that could have arisen under the same circumstances," according to the CDC, contributing to random variation over time or across space. The CDC uses 20 events, corresponding to an RSE of 23%, as the reliability threshold for rates, proportions, and simple ratios from the National Vital Statistics System. The National Health Interview Survey suppresses data if the denominator is based on fewer than 50 cases or the RSE exceeds 30%.[1]

We chose to use very conservative thresholds for determining reliability. For ACS data, we set the threshold at 21% RSE.[2] For indicators based on birth data (Infant Health, Births to Teens, and Prenatal Care), we consider these indicators unreliable in tracts with fewer than 25 births. For indicators based on jobs data (Job availability, Job Growth, and Job Quality), the reliability threshold is 100 jobs; in tracts with fewer than 100 jobs in the 5-mile radius around the tract center, these indicators are considered unreliable. Similarly, for the Business Growth indicator, we used a threshold of 50 employers in the 5-mile buffer around the tract. Indicator values that

[1] http://www.cdc.gov/nchs/data/statnt/statnt24.pdf

[2] We actually use the Relative Margin of Error, or RMOE, to measure reliability, rather than the RSE. Margin of error is just the standard error multiplied by a factor that corresponds to the width of the desired confidence internal. In the case of ACS, the published MOEs correspond to a 90% confidence interval and a multiplier of 1.645. Then RMOE = RSE * 1.645, or RSE = RMOE / 1.645. We use RMOE = 35% as the reliability threshold for ACS data, which corresponds to RSE = 35 / 1.645, or about 21%.

do not meet reliability criteria thresholds are flagged, and the indicators are not used in index calculation in those tracts. The indicator values are still mapped but are displayed with an asterisk to indicate that the values should be interpreted with caution in those tracts.

Furthermore, we did not calculate the index in tracts with a total population less than 100, fewer than 100 households, or where the number of adults incarcerated in the tract exceeded 25% of the citizen voting-age population (CVAP).[3] Tracts that meet these criteria (low population, few households, high percentage of prisoners) are flagged in the data (pop_flag = 1) and appear grey on the index and relevant domain maps.

Variable Transformations

Most of the ROI indicators are percentages, but some are rates or use other metrics. To make it easier to compare indicators, and to ensure that indicators contributed equally to the index, the variables were transformed into a uniform metric.

z-Scores

To facilitate comparisons of different indicators, we transformed standardized each indicator by calculating its z-score. The z-score is a measure of distance from the mean, in this case the mean of all tracts in the region. To calculate the z-score for a particular tract, the value of the indicator in that tract is subtracted from the regional mean and divided by the standard deviation of the indicator, which is a measure of how spread out the indicator values are. A positive z-score indicates the value is above the mean and a negative value indicates the value is below the mean. The farther from the mean, the larger the magnitude of the z-score.

The popups that display in a bar chart all the indicators for a given domain use the z-scores to create the bars. The vertical line represents the mean for the region. Indicators with values below the mean are represented by red bars that extend to the left of the vertical line, while indicators that have values above their means are represented by green bars that extend to the right of the vertical line. The size of the bar is determined by the magnitude of the z-score.

Min-Max Scaling

To construct an index using the geometric mean, it is important that the indicators be positive numbers greater than zero, which is not the case for z-scores. Instead, we use a method called min-max scaling that expresses variables as a percentage of the range between the minimum and maximum values of that variable. For example, if a variable has values ranging from 0 to 10, and one instance of that variable has a value of 1, the transformed version of that variable is 1 / (10 – 0) * 100 = 10%. In other words, the value of 1 is at that 10th percentile of the observed

[3] Prison population was obtained from the 2010 Census Bureau, Table P42: GROUP QUARTERS POPULATION BY GROUP QUARTERS TYPE. CVAP estimates were obtained from a special tabulation of ACS 2007-11, which can be downloaded from http://www.census.gov/rdo/data/voting_age_population_by_citizenship_and_race_cvap.html. The percentage is calculated by dividing the incarcerated population by the total number of United States citizens 18 years of age or older, then multiplying by 100.

values. Even though many indicators were expressed as percentages in their raw form, they were also transformed using this method so that they represented the relative, not the absolute, value of that variable. For example, if the observed range for a variable is 0 – 50%, then a value of 10% would have a value of 20% after transformation: 10 / (50 – 0) * 100 = 20%.

To avoid values of 0, which would result in index scores of 0 (see discussion of the geometric mean below), we set a minimum value of 1 for the transformed indicators. While for most indicators higher values represent more opportunity, there are a few indicators for which this not the case. These indicators were inverted before including them in the index; when this was done it is noted in the indicator descriptions above. For indicators expressed as percentages (teen birth, truancy rate, discipline rate), the inverse is 100 – the percentage. Indicators expressed in other terms and which have no theoretical maximum value (PM2.5 and Years of Potential Life Lost) were inverted by taking the reciprocal of the indicator (1 divided by the indicator).

Index Calculation

An index is a measure which combines and summarizes information from multiple data points. The most common technique for combining multiple measures in an index is to calculate the average, or arithmetic mean, by summing the indicators and dividing the total by the number of indicators. However, indices constructed in this way suffer from the problem of compensability – a high value on one indicator can offset a low score on another. If the indicators are truly interchangeable, this is not an issue, but if a high level of opportunity requires high values on all (or most) indicators, then compensability undermines the ability of the index to capture relative degrees of opportunity across census tracts. For this reason, we chose to use the geometric mean, which penalizes places that have low scores on some indicators.

Geometric Mean

The geometric mean is defined mathematically as the n^{th} root of the product of n values. To take the simplest example, the geometric mean of two values is calculated by multiplying the two values, then taking the square root of the product. For three values, the geometric mean is the third root of the product of the three values, and so on.

The chart below provides an example of how the arithmetic mean and the geometric mean compare for a very simple set of data. The arithmetic mean is the same for every tract, because the two variables have the same sum despite the wide disparities between them. The low values for Variable 1 in Tracts B and C are not offset by higher values for Variable 2, leading to lower values for the geometric mean in these tracts. Note the very low value for the geometric mean in Tract C, despite the very high value on Variable 2.

	Variable 1	Variable 2	Arithmetic Mean	Geometric Mean
Tract A	50	50	50	50
Tract B	10	90	50	30

Tract C	1	99	50	9.95

Domain Means and Index Values

Domain means are calculated by taking the geometric mean of all the indicators in the domain. If any one indicator is either missing or considered unreliable, the domain mean is still calculated over the remaining indicators, taking the $(n-1)^{th}$ root of the product of the n-1 indicators. If more than one indicator is missing, the domain mean is ***not*** calculated, and will have a missing value.

The geometric mean is also used to calculate the overall People and Place indices. The People Index is comprised of six domains, while the Place index is comprised of five domains (we could not find indicators in the Mobility/Transportation domain that met our selection criteria).The People Index is calculated by multiplying the six People domain means, then taking the 6^{th} root of the product. The Place index is calculated by multiplying the five Place domain means, and taking the 5^{th} root of the product. All domain means must be present (no missing values) for the relevant index to be calculated.

How to Interpret Maps and Popup Bar Charts

The ROI website has the ability to map each indicator, each Domain mean, and both People and Place indices. Regardless of which of these items is being mapped, the color scheme is the same. Tracts that are colored green are in the top 20% of tracts in terms of opportunity, light green tracts are in the next lowest 20%, and tracts in yellow are in the middle 20% of all tracts. Tracts that are orange are just below the middle 20%, and tracts that are red are in the bottom 20% of all tracts in the region. The five colors correspond to the five quintiles, or groups of tracts, and the values displayed in the legend represent the actual values for the 20% of tracts in each group. The values for each group depend on the actual distribution of the data for the item being mapped, and will vary from indicator to indicator and domain to domain.

For most indicators, higher values correspond to more opportunity, while for some, higher values correspond to less opportunity. This is the case for two of the three indicators in the Health/Environment: People domain – Years of Life Lost and Teen Birth. You'll notice on maps of these indicators that the legend is "upside down," with red (larger values) at the top and green (smaller values) at the bottom. Tracts in which there is a higher percentage of teen births are not doing as well on this indicator as tracts with a lower percentage of teen births. The other indicator in this domain, Infant Health, or the percent of singleton births that are healthy weight, is measured in a positive direction. For this indicator, like most others, higher values indicate more opportunity. The legend for this indicator's map shows that higher green tracts have higher values.

As noted above, for some indicators, domain, and indices, a tract may be greyed out on the map, or may have an asterisk in it. Grey indicates the tract is missing data, while an asterisk indicates that the indicator has low reliability in that tract (if an indicator is being mapped) or

that the domain mean was calculated with one fewer than the specified number of indicators (if a domain mean is being mapped). Tracts missing more than one indicator in a domain will be displayed in grey on the domain map, and tracts missing on any domain mean will be displayed in grey on the index map.

If you map the index, you can click on a tract to see how the tract fares on the domains that make up the index. Likewise, if you map a domain mean, you can click on any tract in the region to see how the tract fares on the indicators which make up the domain. A popup bar chart will appear showing the tract's values on the underlying indicators (or domains, if viewing an index map) in reference to the regional means. For example, the popup bar chart that appears when the Health/Environment: People domain is the current map shows the three indicators in this domain: Years of Life Lost, Births to Teens, and Infant Health. The vertical line represents the regional mean. Indicators with values below the mean are represented by red bars that extend to the left of the vertical line, while indicators that have values above their means are represented by green bars that extend to the right of the vertical line. The size of the bar is determined by the magnitude of the z-score.

In rare instances, you may find that a tract appears to be above or below based on the domain map, only to find that the popup bar chart shows what appears to be the opposite based on its relative performance on the underlying indicators. This most commonly occurs for the Housing Opportunity: Place domain, which consists of two indicators, Housing Affordability and Housing Adequacy. Housing affordability is the ratio of the median income in the tract to the median home value, while Housing Adequacy is the percentage of housing units in the tract that have at least one room per occupant. In a few tracts, the domain map shows a tract in green, indicating it is in the top 20% in terms of Housing Opportunity: Place, but the popup bar charts displays both indicators in red, indicating below average performance in terms of Housing Adequacy and Housing Affordability. The reverse can also happen: the Housing Opportunity: Place domain score is below average, but the indicators are above average. This anomaly can occur for two reasons. The first is if one of the indicators in the domain is not used in the domain mean calcualtion due to reliability concerns. Then the geometric mean is just the value of the remaining indicator, which may be above average when compared to other tracts, but the domain mean is below average because the tract cannot benefit from the synergistic effect of having two positive indicators. The other is that we use min-max transformed indicators in the calculation, multiplying them together under the geometric mean calculation, but display the z-scores in the popup charts. The anomaly sometimes occurs when indicators are close to the mean and therefore have z-scores that are close to 0, but because of the synergistic effect of multiplying the transformed indicators, the domain mean can be slightly above or below average although the indicators show the opposite relationship to their regional means as evidenced by the popup charts.

Limitations

As with any index, the ROI presents multiple limitations to consider when interpreting and using results. To learn about these, please read the following section carefully.

Indicators

The index lacks key indicators of opportunity because data were unavailable, or unavailable at an appropriate geographic unit. For example, we were unable to identify data sources for indicators in the Mobility/Transportation: Place domain that met our selection criteria. In other domains, we identified, and rejected, indicators we would have liked to include were it not for low reliability at the census tract level.

Weighting

Indices are sensitive to the items included, how they are grouped, and how they are weighted in calculations of summary scores. The 'stepwise equal weighting' method used here ultimately does result in unequal contributions of indicators to the overall index due to differing numbers of indicators within a domain. Research suggests that unless there is a strong rationale for using other weighting schemes, equal weighting of domains and indicators within domains is a reasonable approach. Other weighting schemes may result in different results.

Reliability

As noted above, reliability of some indicators in some tracts is an issue. We chose to display data for these tracts on the maps, highlighting the reliability concerns with an asterisk, but decided not to use them in the index calculation. The problem of reliability is especially acute in tracts with very small populations. We decided not to calculate the index in tracts with fewer than 100 individuals or fewer than 100 households. These tracts are displayed in grey on all maps. Nor do we calculate the index in tracts which have large prison populations (over 25% of the estimated citizen voting age population), as estimates for these tracts may be skewed by the presence of large numbers of prisoners.

Validity

Assessing the validity of a multidimensional index is difficult because it is unclear what criteria to use. Going forward, it will be important to compare this index to others, assess its ability to predict various concurrent and future outcomes, and assess its utility as a tool for policy-makers, community development practitioners and advocates.

All limitations should be considered when using this index. Despite these limitations, the ROI captures many key aspects of regional opportunity, including factors that are vital for the development of healthy, sustainable communities. The ROI allows users to explore relationships between these factors, providing insight into the distribution of opportunity structures in California's Central Valley and helping identify places in need of further investment.