haas institute
FOR A FAIR AND INCLUSIVE SOCIETY

Methodology
2018

Mapping Opportunity

A subtitle to be determined

Authors:

This brief is published by the Haas Institute for a Fair and Inclusive Society at UC Berkeley

About the Authors

Phuong

Samir

Heather

Stephen

Additional research assistance was provided by Heather Bromfield, James, Arthur Gailes

Reviewers

Samir Gambhir

Stephen Menendian

Arthur Gailes

XXXXX XXXXX

XXXX XXXXX

Copyeditor

Marc Abizeid

Design & Layout

Rachelle P

Charts

Samir Gambhir

Phuong Tseng

Citation

Nicole Montojo, Stephen Barton, Eli Moore. "Opening the Door for Rent Control: Toward a Comprehensive Approach to Protecting California's Renters" Haas Institute for a Fair and Inclusive Society, University of California, Berkeley, CA. September 2018.

On the Cover
A landscape photograph showing rows of houses set in front of a mountain range

Contact

460 Stephens Hall
Berkeley, CA 94720-2330
Tel 510-642-3326
haasinstitute.berkeley.edu

Acknowledgements

We thank our peers for their work that helped inform this research, including:

Program

Contents

Introduction

Neighborhood conditions and access to opportunity play a major role in improving the life chances of its residents. Social science research has highlighted how critical housing location is in creating positive life outcomes for residents.[1] Opportunity structures, such as access to good schools and employment, define neighborhoods as opportunity-rich or opportunity-poor.[2] Access or barriers to opportunity structures can enhance or inhibit a person's chances of upward economic mobility. It impacts not only a person's life chances but can be instrumental in intergenerational economic mobility. The goal of the opportunity mapping project is to assess racial inequity in space, identify institutions and structures that establish or perpetuate inequity in access to opportunity, and provide policy recommendations to remove barriers to opportunity for all.

Opportunity mapping quantifies cumulative effects of neighborhood conditions that pose barriers to structures of opportunity, and to address fair housing issues through community development. Our approach to mapping opportunity across space employs quantitative methods and Geographic Information Systems (GIS) software in an attempt to quantify and visualize this abstract phenomenon. The domains under which these structural barriers are analyzed are housing, education, economy, and public safety. Multiple indicators have been identified under each of these domains as proxy for opportunity. This document describes our indicators, data sources and our methodology in developing this tool.

METHODOLOGY

I. Identifying Indicators

The process of selecting indicators involved a deep review of social science literature to understand which data sources can serve as proxies for opportunity at the neighborhood level. Though neighborhood could be defined in many ways, we use census tracts as proxies for neighborhoods. Therefore, we searched for data at and apportioned data to the census tract level, wherever possible.

For each region's Opportunity Index, we categorize our indicators into four main domains of opportunity: education, economic and mobility, housing and neighborhoods, and conduit.[3] Within these four domains, we compiled nineteen indicators at the census tract level from various federal, state, and local data sources.[4]

1 john a. powell, Remedial Phase Expert Report of john powell in Thompson V. HUD (Columbus, OH: Kirwan Institute for the Study of Race and Ethnicity, 2005), http://www.kirwaninstitute.osu.edu/reports/2005/09_2005_ThompsonvHUDRemedialReport.pdf.

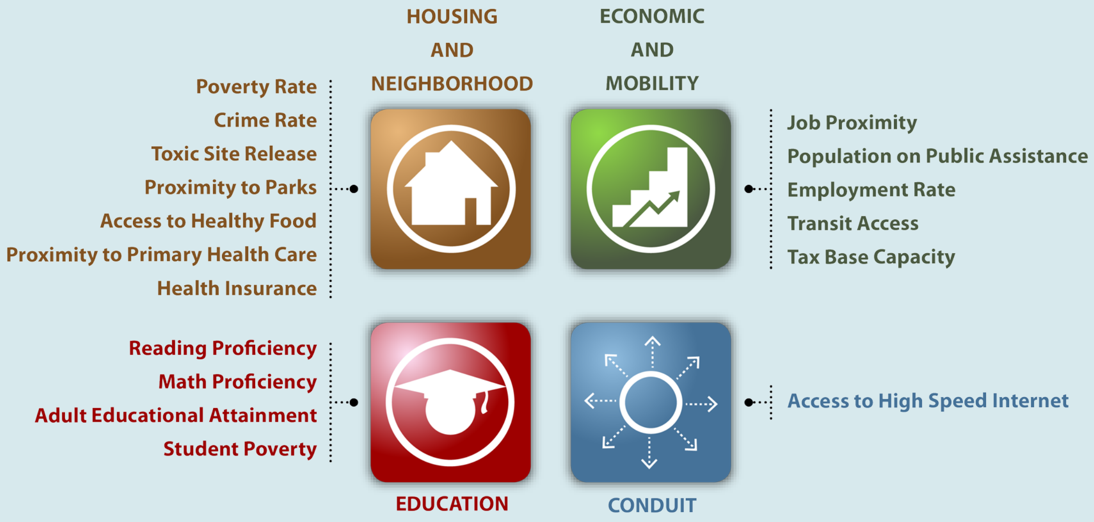
2 George C. Galster, "Urban Opportunity Structure and Racial/Ethnic Polarization" (College of Urban, Labor, and Metropolitan Affairs, Wayne State University, 2005), https://pdfs.semanticscholar.org/4edb/eaff059c9472a6847dc6b4b61e7931cd97c0.pdf.

3 Other indices of opportunity often categorize their indicators into different kinds of domains, especially if they utilize multiple indicators to measure opportunity. Depending on the methodology and purpose of these indices, the number of indicators and domains vary from index to index.

4 Data sources are from the United States Environmental Protection Agency (EPA), National Center for Education Statistics (NCES), Tetrad's US Demographics crime risk index drawn from the FBI Uniform Crime Reporting database, ESRI, and American Community Surveys (ACS) from the United States Census Bureau.

Domains and Indicators

This table displays a list of opportunity indicators by education, economic and mobility, and housing and neighborhoods domains, and conduit.



II. Indicator Selection, Measures, and Data Collection

The indicator selection process requires a thorough and comprehensive approach to select data sources that contribute to, or detract from, neighborhood opportunity. In addition, we relied on the existing social science literature to inform decision-making on the measure for each indicator and sought measurements that would be equally viable in each region of the country. As previously stated, we used the census tract as our unit of analysis. Therefore, we searched for data, that are available and reliable, at the census tract level, wherever possible. To ensure the accuracy and currency of our data, we relied on data from the U.S. Census Bureau and other reputable data sources.

METHODOLOGY DESIGN

While certain indicators have fairly straightforward measurements (for example, the unemployment rate is calculated for each census tract by the American Community Surveys and is therefore straightforward to operationalize), some indicators represent abstract ideas and thus require deeper consideration before the most appropriate measurement can be chosen. For example, while there is a consensus that reading and math proficiency scores demonstrate the performance of K-12 schools, determining which data points are appropriate measurements of reading and math scores is less clear. There are options to use scores from different grade levels, and different data sources.

Data for some indicators, such as neighborhood schools and toxic sites, were available at a point location, and had to be transformed using GIS procedures to assign the data to one or many tracts that are impacted by or contribute to the outcomes. Some of the GIS procedures employed include buffers, Voronoi Polygons, area aggregation, geocoding, and geostatistical analysis tools.

All efforts were made to find any missing values from the data. Though our intention was to minimize missing data values while ensuring the quality of the data, we reviewed and assessed each dataset independently to fill in the missing data. In some case, alternative sources, such as using the most recent year of available data (in lieu of the current year's data), were utilized to fill in the gaps. Other strategies involved replacing each missing value with the mean value for the indicator.

III. Standardizing Indicators and Domains

The conceptual framework of this project allows us to include indicators reflecting various aspects and outcomes of life. These indicators often have different units of measurement. For example, income data is measured in dollars, poverty rate is measured by percentage, and toxic release by pounds. In order to make these measurements comparable for each tract, we standardized the values by utilizing a statistical measure called z-score.[5] This method normalizes our data for each indicator (using standard deviations) and renders all indicators unit-less while assessing how far each observation is from the mean.

Computed z-score values for any dataset inform us about the directionality of each data point. If a census tract's indicator value has a negative z-score, then the observation is less than the mean, while a positive z-score signifies that the observation is greater than the mean. Similarly, a negative z-score implies a negative impact on overall levels of opportunity while a positive z-score suggests a positive impact. Where necessary, we adjusted the directionality of the z-score values where the directionality does not automatically correlate with how the indicator affects opportunity. For example, for our "proximity to toxic waste sites" indicator, the further away that toxic waste sites are from a neighborhood, the higher the level of opportunity for residents, and thus the z-score for this indicator does not need to be adjusted (larger distances indicate positive outcomes). By contrast, for

5 The formula used for calculating the z-score is the following: $(n-\mu)\div\sigma$, where n is the observation, μ is the mean, and σ is the standard deviation.

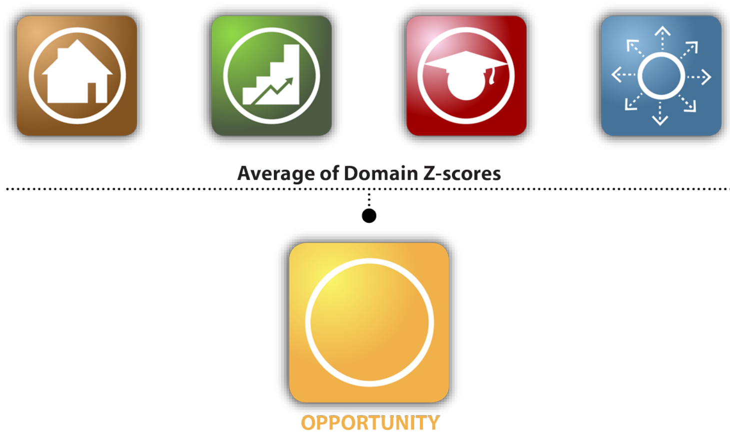
our "proximity to parks and open spaces" indicator, the farther away parks are from a neighborhood, the lower the level of opportunity for residents. We multiply the z-scores by -1 so that negative signs indicate larger distances and reduced opportunity.

IV. Computing the Opportunity Index

As a first step toward computing the overall opportunity index, each indicator's z-score is calculated. For each census tract, the domain z-score is the average of z-scores for all indicators within the domain. The overall opportunity scores are computed by summing the z-scores of all three domains and dividing it by the total number of domains. In other words, the opportunity score for each census tract is the average z-score of all domains. We subsequently apply three filters to all the census tracts in the region (discussed in section A, below) and finally we sort the index from the highest to the lowest opportunity score and distribute the tracts into our five categories of opportunity—Very High, High, Moderate, Low, and Very Low—using the procedure discussed in Section B, below.

A. Averaging Opportunity Domains

Census tracts that initially appear to have a high opportunity score may in fact have other characteristics which are proven in academic literature to overwhelmingly reduce opportunities for people living in these neighborhoods, even when other positive neighborhood attributes are present. For this reason,



METHODOLOGY PART TWO

we have chosen to use "filters" to separate out tracts with specific characteristics: high levels of racial segregation, a high percentage of single-parent households, and high concentration of poverty.[6] We utilized specific thresholds for each as described below.

For our calculation of racial segregation, we use Divergence Index, a metric which compares race proportions at the tract and county levels. For example, the metric compares the percentage of white people at the tract level to white people at the county level, Black people at the tract level to Black people at the county level, and so on for each of our five selected racial categories (white, Black, Asian, Hispanic, and the sum of people in other racial categories). The Divergence Index is calculated using these comparative percentages between tracts and county. The higher the index value, the higher the level of segregation.[7,8] In our methodology, we categorize the tracts for each county into three categories with an equal number of records (quintiles or relative ranking). These three categories are high segregation, moderate segregation, and low segregation. Thus, the higher the value entails a larger divergence and a high level of segregation.

We have refined these thresholds to help us capture the extent of our filters. Ultimately, the institute workgroup decided to use these two criteria and threshold values:[9]

1) Single-parent families >= 30 percent

and

Poverty rate >= 30 percent,

OR

2) Poverty rate >= 30 percent

and

High segregation in the Divergence Index

In other words, if a tract has 30 percent or more of the population living in a single-parent household and falls under the federal poverty line, it will be captured by our first criteria. A tract will also be captured if it is highly segregated and has 30 percent or more of the population falling under the 200 percent federal poverty line. There is a total of XXX census tracts captured by these two criteria.

6 Raj Chetty, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner, "The equality of opportunity project," http://www.equality-of-opportunity.org.

7 Elizabeth Roberto and Jackelyn Hwang, "Barriers to Integration: Physical Boundaries and the Spatial Structure of Residential Segregation" (working paper, Cornell University, 2015), https://arxiv.org/abs/1509.02574.

8 Stephen Menendian and Samir Gambhir, Racial Segregation in the San Francisco Bay Area: Part 1 (Berkeley, CA: Haas Institute for a Fair and Inclusive Society. 2018), accessed November 27, 2018, https://haasinstitute.berkeley.edu/racial-segregation-san-francisco-bay-area.

9 Prior to finalizing our filter thresholds, we tested and refined the filters based on the following levels: Single-parent families >= 25%, 30%, 35% and 40% (for reference, the national average is 27%, calculated using the American Community Survey 1-year estimate to compute the percentage for the nation). We also experimented with threshold values for the poverty rate >= 20%, 30%, 35%, and 40%, and for entropy >= 0.10, 0.12, 0.15, 0.17 - 0.20, 0.6, and 0.7.

B. Procedure to Allocate Filtered & Unfiltered Tracts

A subset of census tracts that are filtered according to the above procedure are automatically assigned to the bottom two opportunity categories. Half of these filtered tracts are assigned to the "Very Low" category and half are assigned to the "Low" category, based on their opportunity index values. The total number of tracts which are filtered varies by region according to the demographic features of the respective census tracts, and thus these filters serve as an absolute, rather than relative, measure of opportunity.

The remaining tracts in our study area—those which were not filtered out—are reclassified as "Moderate," "High," or "Very High" based on their respective index values. Thus, 20 percent of the unfiltered tracts with the highest index values are assigned to the "Very High" category, the next highest 20 percent to the "High" category, and the remaining are assigned to the "Moderate" category.

V: Limitations

The workgroup recognizes that there are several limitations to the computation of this index due to data availability and several methodological choices. One of these limitations is the weighting of indicators. Although we do not weight the indicators within the index, the indicators unintentionally receive variable weights because of the process by which we average our domains. Due to the differences in the numbers of indicators per domain and the fact that each domain is weighted equally in the final opportunity index, indicators in domains with fewer variables receive a greater weight than indicators in domains with more variables or elements. For example, educational opportunity has a total of five indicators; therefore, each indicator contributes to one-fifth of the overall domain score, while each indicator in the housing and neighborhood opportunity domain (which has eight indicators) contributes one-eighth to the total domain score. However, arranging indicators by domains and averaging them to compute the final z-scores prevents one domain with more indicators from contributing more to the overall index because doing so weights each domain one-fourth of the overall index score.

To handle missing data values, when we assigned a z-score value of "0" to missing entries, we are suggesting that all missing data values are equal to the mean since z-scores of zero normally indicate that the observation is 0 standard deviations from the average value. This assumption could impact the index value.

V. Ongoing Efforts

Some of our ongoing efforts are to refine our method for understanding and calculating tax capacity; measuring commuting time and single-parent families; identifying an appropriate threshold for high segregation; and heteroskedasticity and collinearity in data analysis. The current metrics provide a snapshot of opportunity, and thus, one of our priorities in the future iterations is to develop a change overtime methodology.

One of the challenges we face in this iteration is the incorporation of an economic measure tax capacity. As of now, it is measured by attributing municipal revenue per capita to each census tract that falls within the municipal boundary to quantify the extent of opportunity hoarding at the municipal-level.[10] This method has its limitations because it assumes that every resident in the neighborhood contributes an equal amount of taxes. Yet, in reality, a municipality consists of residents who fall into a range of tax brackets, and thus, residents in the higher tax bracket might contribute more while residents in the lower tax bracket might contribute less in taxes. Furthermore, residents in the higher income tax bracket can also take advantages of the other tax-deferred or tax-free benefits as a way to lower their taxes. Not only is the difference in contribution at the individual level vital to the overall revenue generated at the municipal-level, the variation in revenue generated by each municipality provides insights into which municipality

10 Richard V. Reeves, Dream Hoarders: How the American Upper Middle Class Is Leaving Everyone Else in the Dust, Why That Is a Problem, and What to Do about It (Washington, D.C.: Brookings Institution Press, 2018).

has a higher tax base versus which municipality has a lower tax base. The higher the tax base means more revenue is generated for that municipality.

Another limitation that is important to discuss in this section is the fact that half of our dataset were gathered from the American Community Survey's five-year estimates dataset, which are averaged over the years 2013-2017, and did not consider the margins of error (MOE) in each census tract before computing the domain and the overall index. Using the census tract as our level of analysis, a geography with small populations, may mean there are high percentage MOEs in the estimates which we use to perform our calculations, particularly when sub-populations are considered (such as with race data). Not considering the impact of MOE in our calculation may impact the accuracy of our analysis.

PLACEHOLDER: NARRATIVE FOR COMMUTE TIME

PLACEHOLDER FOR TEXT AND FINAL OPPORTUNITY MAPS

The Haas Institute for a Fair and Inclusive Society brings together researchers, community stakeholders, and policymakers to identify and challenge the barriers to an inclusive, just, and sustainable society in order to create transformative change.

Contact

460 Stephens Hall
Berkeley, CA 94720-2330
Tel 510-642-3326
haasinstitute.berkeley.edu

@haasinstitute