Documentation: Neighborhood Affordable Housing Priority Interactive Tool

Living in a neighborhood with low poverty and crime, access to transit and jobs, and a racially diverse population confers a range of benefits, big and small. Better transit enables residents to access and keep their jobs for the long term. Well-resourced neighborhood schools empower children to perform at their best. Residents report feeling happier, safer, and healthier. Typical evidence-based metrics of neighborhood-level opportunity include neighborhood safety, concentration of entry-level jobs, transit access, and school quality. Yet low-income residents can rarely afford to rent or buy homes in neighborhoods with these characteristics.

The District acknowledges both the desirability and the need to subsidize affordable housing development in neighborhoods with access to more opportunities. But subsidizing housing in those neighborhoods, where market competition for land is fierce, is costlier. Maximizing the impact of DC’s affordable housing funds is critical because those resources, while substantial, are nevertheless dwarfed by need.

This interactive tool allows the user to attach different levels of importance to five different neighborhood indices related to opportunity: jobs proximity, poverty, racial diversity, and violent crime, as well as the cost to develop housing, as measured by land value.

This project aims to provide a tool for policymakers and affordable housing providers to identify neighborhoods with characteristics that promote economic opportunity and well-being, and to weigh the cost of subsidizing housing in those areas. Ideally, the tool will reveal areas that score high on various opportunity metrics, yet are relatively lower in cost than other, similarly ranked areas.

**Jobs Proximity Index**

**Background**

In cities throughout the country, low-income residents and people of color often live in the city center, while many job opportunities are dispersed throughout the city and metropolitan region in places not sufficiently accessible by transit. Black residents are particularly impacted by this mismatch between home and work, which may contribute to higher levels of unemployment and lower earnings. Improved access to public transit is a concrete way to address the problem of job access, as public transit access increases an individual’s likelihood of retaining employment.

The jobs proximity index measures a census tract’s access to jobs in DC and the surrounding area via public transit. Tracts receive a higher score if they have lower transit times, more job opportunities, and less competition for those jobs. Conversely, tracts receive a lower score if they have higher transit times, fewer job opportunities, and more competition for those jobs. Possible scores range from 0 to 100, where 0 is the worst and 100 is the best in terms of job proximity.

Measures of job access can be affected by the size of the study area. Selecting a study area that is too small can artificially reduce the number of available jobs; the reverse applies with a study area that is too large.
Since the study area must correspond to people’s actual employment patterns and job-seeking behavior, we first identified the area within which lower-income DC residents typically work, using American Community Survey PUMS Microdata.

We found that 79 percent of lower-income (below 60 percent of the median income) DC workers worked either within DC, or in a jurisdiction immediately surrounding DC.\(^6\) The jobs proximity study area was thus defined as census tracts in DC or in those jurisdictions: District of Columbia; in Virginia, Arlington County, Fairfax County, Alexandria city, Fairfax city, Falls Church city, Manassas city, and Manassas Park city; and in Maryland, Montgomery County and Prince George’s County.

**Data**

Public transit time data was obtained using the Google Maps Distance Matrix API. Transit times were calculated from the centroid of every DC Census tract to the centroids of every other census tract in the study area.

Jobs data were obtained from the Census Bureau’s Longitudinal Employer-Household Dynamics (LEHD), specifically 2014 Workplace Area Characteristics (WAC) data from LEHD Origin-Destination Employment Statistics.\(^7\) These data provide jobs information at the block group level and were aggregated to the census tract level.

Labor supply data came from 2015 ACS 5-year estimates on the civilian labor force (table DP03). Field HC01_VC11 served as the estimated number of residents in the labor force.

**Formula**

The jobs proximity index relates positively weights each destination tract for its job opportunities and negatively weights the tract for its competition. It is a distance decay model where farther tracts receive less importance than nearer tracts. The jobs proximity index is calculated as follows:\(^8\)

\[
A_i = \frac{\sum_{j=1}^{n} \frac{E_j}{d_{ij}^2}}{\sum_{j=1}^{n} \frac{L_j}{d_{ij}^2}}
\]

Where \(i\) is the origin tract, \(j\) is the destination tract, \(E\) is the number of available jobs in the tract, \(L\) is the supply of workers living in the tract, and \(d\) is the distance, in minutes, separating the two tracts.

The tract that receives the most importance is the tract where \(i\) and \(j\) are the same, since that is the tract where one lives. In the Google Maps API distance matrix, cases where the origin and the destination were the same computed a zero distance. Zero distances were programatically replaced with a transit time of 1 to receive the most importance in the final jobs proximity index. To maintain consistency, all transit times were calculated with a forced arrival time of 9 a.m. on Wednesday, June 21, 2017.

Some areas were determined to be impossible to access using public transit by 9 a.m. on a weekday. In these cases, empty distance values were returned as a result. Whether one leaves these values empty or replaces these values with an infinite distance, the contributions of \(E\) and \(L\) are canceled out by their respective
denominators. Therefore, areas that cannot be accessed by public transit contribute nothing to the final summation and may be ignored.

Once $A_i$ was calculated for every origin tract, these values were converted to tract percentile ranking by employing the empirical cumulative distribution function. The final index ranges between possible values of 0 and 100, where 100 is ideal.

**Neighborhood Poverty Index**

**Background**

The impact of moving to low-poverty neighborhoods has been extensively studied. In 1993, the federal government began a large-scale experiment called Moving to Opportunity (MTO) that facilitated participants’ moves from census tracts with high poverty rates to census tracts with poverty rates below 10 percent. Analysis of MTO outcomes identified positive impacts on participants’ educational attainment, health, and economic well-being. People who relocate to low-poverty neighborhoods report that they are generally happier and less stressed. There are long-term impacts as well: children who moved to low-poverty neighborhoods at a young age earn an additional $302,000 over their lifetimes, on average, than if they had not moved. While poverty cannot tell the entire story, it remains an important variable to consider.

The neighborhood poverty index measures the estimated percentage of households living below the poverty level. Tracts with lower estimated poverty rates receive a higher score; tracts with higher poverty rates receive a lower score. Possible scores range from 0 to 100, where 100 has the lowest estimated incidence of poverty.

**Data**

Data on poverty rates were obtained from 2015 ACS 5-year estimates on residents living below the poverty level (table B17017) for DC census tracts. Field HD01_VD02 served as the estimated number of residents with income below the poverty level. Field HD01_VD01 served as the estimated total number of residents in each tract. Estimates of the percentage of residents living below the poverty level in each tract were obtained by dividing field HD01_VD02 by field HD01_VD01.

**Formula**

- Census tracts with estimated poverty levels below 10 percent received a score of 100.
- Census tracts with estimated poverty levels between 10 percent and 40 percent received a score of 67.
- Census tracts with estimated poverty levels above 40 percent received a score of 33.
- Census tracts with populations under 100 received a score of 0, because these small populations tend to have highly skewed estimates.

**Residential Land Cost Index**

**Background**

The cost of subsidizing housing in a neighborhood is necessarily an important piece of this tool. While DC government has recently begun to award funding priority points to affordable housing developments in high-cost areas, at the same time, it prioritizes affordable housing developments that cost the District less overall.
While construction costs (labor and commodities) are relatively consistent across neighborhoods, land costs are not. In neighborhoods with high demand for market-rate housing or mixed-use development, competition for land can be fierce, leading to high sale prices.

The residential land cost index measures the appraised value of land in DC. Possible scores range from 0 to 100, where 0 is the most expensive and 100 is the least expensive.

Data
The DC Office of Tax and Revenue (OTR) conducts an annual assessment of property values in the District. According to OTR documents, the residential cost model by which properties are assessed is calibrated using a non-linear regression model developed from citywide market analysis of qualified sales. One element of the cost model is a residential base lot value for each assessment neighborhood. The base lot value was selected as the basis for this tool’s land cost index because it reflects both the size of a typical lot and the cost per square foot of the lot—capturing the typical per-lot cost.

It is important to note that this index does not account for the variation in land ownership turnover across neighborhoods. For instance, whereas one neighborhood may be dominated by long-time owners or uses, with few properties put up for sale in any given year, another neighborhood may have more sales, presenting more chances to buy parcels for the purpose of developing affordable housing.

Formula
The DC Office of Tax and Revenue (OTR) computes base rates for property values over special neighborhoods and subneighborhoods specific to their office. Property tax rates were appended to subneighborhoods when possible, and to neighborhoods when subneighborhoods were not defined. Since the OTR’s Real Property Tax Administration neighborhoods do not conform to census tract boundaries, OTR’s data was spatially transformed to match census tracts.

Areas containing public land, such as Rock Creek Park and the National Arboretum, are not assessed taxes and therefore have base lot values of zero. Converting from OTR neighborhoods and subneighborhoods to census tracts without adjusting for zero base lot values results in the devaluation of all census tracts containing public land. To correct for the bias, we computed the centroid of every OTR neighborhood and subneighborhood along with the planar distance between every centroid. Neighborhoods and subneighborhoods with a base lot value of zero were assigned the base lot value of the nearest neighbor with a base lot value above zero.

Converting spatial units from OTR neighborhoods and subneighborhoods to census tract boundaries required four steps. First, base lot values from OTR was converted to raster format with cells of 330 feet by 330 feet. Second, census tract identification numbers were converted to a raster file with the same resolution. Third, base lot value data was appended to census tract information by using a zonal summation, where the zone was the census tract identification number and the summation was the sum of all base lot values in the new zone. Fourth, the zonal summation was divided by the number of cells in the zone to determine the base lot value for the census tract.

To obtain the final metric, the inverse base lot value was normalized using the formula:

\[ idx = \frac{v - \bar{v}}{s_v} \]
Where $v = \frac{1}{b}$ and $b$ is the base lot value for each census tract. The normalized inverse base lot values were then converted to tract percentile ranking by employing the empirical cumulative distribution function. Possible scores range from 0 to 100, where 0 is the most expensive and 100 is the least expensive.

**Racial Diversity Index**

**Background**

There are at least four reasons why it is important to consider the racial diversity of DC neighborhoods. First, racial segregation of neighborhoods reflects a history of racist and discriminatory policies. Second, U.S. neighborhoods remain highly segregated in the present: in 2010, “half of the metropolitan black population [lived] in neighborhoods that were home to only 3.6 percent of the nonblack population.” Third, residents of highly segregated neighborhoods have more trouble accessing jobs and information and often face discrimination from employers, due to the historic denial of amenities, services, and wealth-building opportunities to those neighborhoods. Fourth, present inequities tend to replicate in the future. Research indicates that growing up in disadvantaged neighborhoods adversely affects children’s development and educational outcomes, thereby passing on disadvantage to future generations.

The racial diversity index compares the racial diversity of each census tract to the diversity of the District as a whole. Possible scores range from 0 to 100, where 0 is the least diverse tract and 100 is the most diverse tract.

The racial diversity index is only a slight adaptation of the racial dissimilarity index, which is a prominent segregation metric. The racial dissimilarity index can be written as follows:

$$D = \frac{1}{2} \sum_{i=1}^{k} |x_i - y_i|$$

$$x_i = \frac{N_i}{N}$$

$$y_i = \frac{W_i}{W}$$

Where $N_i$ is the number of non-white residents in a census tract, $N$ is the total number of non-white residents in the study area, $W_i$ is the number of whites in a census tract, and $W$ is the total number of white residents in the study area.

When calculated in this way, the racial dissimilarity index determines the proportion of either of the groups that would have to move in order to match the overall racial makeup of the study area. There are two issues with the racial dissimilarity index as written. First, it computes a summation across the study area, thereby eliminating the differences between individual census tracts. Second, it can only compare differences between two groups.

The racial diversity index addresses the first shortcoming: instead of observing the overall segregation of a city, the racial diversity index allows the user to compare differences between individual census tracts. It does not address the second shortcoming: like the racial dissimilarity index, it can only compare differences between two groups.

One issue with the racial diversity index is that it treats tract-level concentration of white residents and of non-white residents as equally disadvantageous. The index was chosen primarily because the traditional measures of racial segregation, the proportion of residents in a tract who are non-white, is highly correlated...
with other metrics in the tool, such as poverty and violent crime, and thus would not change the picture of the overall opportunity index much. (The choice of a diversity index should not be taken to mean that it is not desirable to increase housing opportunities for low-income and people of color in advantaged, majority-white neighborhoods. It is.)

The racial diversity index used 2015 ACS 5-year estimates on race and ethnic origin (table B02001) for DC census tracts. Since the racial diversity index can only measure differences between two groups, this index measured white – non-white difference. Field HD01_VD02 served as the estimate of white residents in DC. The summation of fields HD01_VD03 through HD01_VD08 served as the estimate of non-white residents in DC. Field HD01_VD01 served as the estimate of total residents in DC regardless of race or ethnic origin.

**Formula**

The racial diversity index is the normalized inverse absolute difference between the percentage of white and nonwhite residents in each DC census tract. First, the inverse absolute difference is computed using the formula:

\[
d = \frac{1}{\left| \frac{N_i}{N_T} - \frac{W_i}{W_T} \right|}
\]

Where \(N_i\) is the estimated number of non-white residents in a census tract, \(N_T\) is the estimated number of non-white residents in all census tracts, \(W_i\) is the estimated number of white residents in a census tract, and \(W_T\) is the estimated number of white residents in all census tracts.

\(\frac{W_i}{W_T}\) and \(\frac{N_i}{N_T}\) are simple ratios showing the number of white and non-white residents in the tract relative to the number of white and non-white residents in the District. The absolute difference between these two ratios gives some sense of racial diversity in DC census tracts, since we would expect that large differences connote tracts that are not diverse. We take the inverse of this value because our model prefers diverse tracts.

Second, the inverse absolute difference for each tract is normalized using the formula:

\[
idx = \frac{d - \bar{d}}{s_d}
\]

Possible scores range from 0 to 100, where 0 is the least diverse tract and 100 is the most diverse tract.

**Violent Crime Index**

**Background**

Research indicates that increased exposure to neighborhood violence negatively impacts residents’ health, educational attainment, and economic outcomes. Conversely, living in neighborhoods with less crime can reduce stress levels and promote pro-social behavior.

The violent crime index measures the violent crime rates of each census tract. Possible scores range from 0 to 100, where 0 indicates the highest violent crime rate and 100 indicates the lowest violent crime rate.
DC’s Metropolitan Police Department defines the following types of crime as violent crime: homicide, sex abuse, assault with a dangerous weapon, and robbery.\textsuperscript{25}

**Data**

The violent crime index used data from a point shapefile containing all DC crime incidents in 2015, downloaded from DC’s Open Data Portal.\textsuperscript{26} Violent crime counts were spatially aggregated by census tract and standardized as rates per 1,000 residents.

**Formula**

The crime index is the normalized inverse violent crime rate in each DC census tract. This is:

\[
idx = \frac{c - \bar{c}}{s_c}
\]

Where \(c = \frac{1}{r}\) and \(r\) is the violent crime rate for each census tract. This metric was rescaled to an index with values ranging from 0 to 100. Tracts where zero violent crimes were reported in 2015 were excluded from the computation and given a perfect score of 100.

**Affordable Housing Priority Index**

**Background**

The strength of this affordable housing tool is that it allows users to interactively view the interaction of neighborhood quality metrics alongside the cost of development. The affordable housing priority index is the metric that enables this visualization. The user can weigh the importance of the other five indices — jobs proximity, neighborhood poverty, residential land cost, racial diversity, and violent crime — and dynamically update the prioritization of areas in DC for affordable housing development.

Possible scores of the affordable housing priority index range from 0 to 100, where 0 indicates the least-optimal for affordable housing development and 100 is the most optimal.

**Data**

The affordable housing priority index uses two data sources: first, the user-supplied weights assigned to each of the five indices; and second, the values of the five indices for each tract. Assigned weights can range from 0 to 100, where 100 designates the greatest priority. The values of the indices were calculated using the procedures outlined in this document.

**Formula**

The affordable housing priority index is the weighted average of the jobs proximity index, neighborhood poverty index, residential land cost index, racial diversity index, and violent crime index, where the weights are furnished by the user in the tool. The formula for the index is as follows:

\[
idx = \frac{\sum_{i=1}^{n}(x_iw_i)}{\sum_{i=1}^{n}(w_i)}
\]

Where \(w\) is the weight assigned to each index and \(x\) is the value of each index.
A Note on Schools

At last count, there was a roughly even split in the number of public schools and charter schools in DC. Both types of schools are well-represented in DC, and they both have different ways of measuring school performance. We elected not to include a school quality metric in this study for two reasons: first, because some researchers caution against their use; and second, because the researchers who do use these metrics have wildly different approaches to measuring school quality. Therefore, we provide links to information about DC schools but otherwise allow the user to make their own judgments about school quality in DC.


9 2015 American Community Survey 5-year PUMS and PUMA crosswalk; HUD Income Limits.


12 Ibid.


18 Ibid.


22 See the following for discussions of the dissimilarity index:


30 For a survey of different approaches to school quality metrics, see:


