Philadelphia Playspace Mapping

Quantitative Methods Technical Brief

Outline

- Introduction
- Prioritization Model:
    - Data variables
    - Z-Score Methodology
    - Prioritization Weights
    - Upsampling to Neighborhoods and Planning Districts
- Data Sources

Introduction/Overview:

Overall, the purpose of this map is to develop an analysis of playspace inequity gaps for Philadelphia's youngest children. In close consultation with KABOOM!, we at Child Care Aware® of America (CCAoA) developed an online map using available data to approximate playspace inequity for children ages 0-6, especially in child care settings. We included demographic data, location of existing playspaces and other early learning infrastructure, and location of child care settings. The map features "Priority Score," layers which represent the results of our playspace equity model. These figures help our organizations focus on specific areas throughout the city that may be higher-priority spaces for early learning investments. In addition to these priority data, there are several other map layers included. Some of these layers featured are directly included in the equity model, while others are not but provide additional context about neighborhoods' early learning infrastructure landscapes.

Understanding Prioritization Model

The playspace equity prioritization model is calculated by adding each census tract's weighted z-scores for 24 key socioeconomic and early learning equity variables. These 24 variables are intentionally selected measures of socioeconomic, early learning, and play equity KABOOM! & CCAoA want to consider when deciding where to direct future investments (e.g., playspaces). The map features census tract, neighborhood, and Philadelphia planning district priority levels based on the model's output priority scores. The higher the score, the more priority the model suggests that the area should receive when considering equity-based investments.

Prioritization Model Variables

KABOOM! provided CCAoA socioeconomic data they obtained from the mySidewalk database at the census tract unit of analysis (UOA). These variables included demographic, built environment, and socio-economic data curated from several sources, including the US Census. Examples of these variables include children with low access to healthy food, high unemployment rates, and low life expectancy at birth. CCAoA built upon this data by including additional state and local child care datasets which introduced early care and education infrastructure into the model. Finally, a custom-created, comprehensive collection of Philadelphia playground data were constructed from myriad datasets. Table 1 shows a complete list of the 24 variables included in the prioritization model.

Variable	Notes	Variable Source	Weight*
Minority Population per capita (2016-2020)	Minority Population normalized by total population.	American Community Survey	2.5
Population of children (under 18) per capita (2016-2020)**	Child Population normalized by total population.	American Community Survey	2.5
Average walking distance from child care facility to nearest playground***	Helps identify concentrations of ECE providers without playground access. Calculated by CCAoA using child care providers and walkable streets and paths (from Open Street Map) in a GIS Origin-Destination Matrix Network Analysis.	Playspace Accessibility Model	2.5
Median Household Income (2016-2020)	Median Income of Total Occupied Housing	American Community Survey	-2.5
Percent Area Covered by Playgrounds (2022)***	Shows extent of existing playspace infrastructure. Calculated by CCAoA. Note: this does not account for playspace quality.	Playspace Accessibility Model	-2.5
Potential child care demand per capita (2016 – 2020)**	Share of children with all household parents in the workforce. Standard research proxy for child care demand.	American Community Survey	2
Children with Low Access to Healthy Food within 1/2 Mile per capita (2019)**	Share of children living in food deserts. Proxy for health that focuses on children within a small / well defined geographic area.	USDA ERS Food Access Research Atlas	2
Age 18 and Under with Medicaid or Means-tested Public Coverage per capita (2016-2020)**	Proxy for income, access, and health, all of which are important for prioritization purposes.	American Community Survey	2
Number of Parks (2018)	Proxy for existing access - important to identify potential existing / new areas for playspaces.	openICPSR National Neighborhood Data Archive (NaNDA)	-2
Percent Area Covered by Parks (2018)	Proxy for existing access - important to identify potential existing / new areas for playspaces.	openICPSR National Neighborhood Data Archive (NaNDA)	-2
Life Expectancy at Birth (2010-2015)	Important health outcome. Should prioritize areas with low life expectancy, under assumption that playspaces will contribute to	CDC National Center for Health Statistics	-2

	better physical health, and potentially improved life expectancy.		
Number of Schools	Proxy for existing playspace access - important to identify potential existing / new areas for playspaces.	National Center for Education Statistics	-2
Language Isolation per capita (2016-2020)	Useful as additional proxy / complement to minority population indicator.	American Community Survey	1.5
Children Under 5 with a Disability per capita (2016-2020)**	Important because it focuses on population with additional needs. However, not every one of KABOOM!'s builds serves this population.	American Community Survey	1.5
Vehicles Available for Occupied Housing Unit - No Vehicles per capita (2016-2020)	Assumption: this serves as a proxy for access - if households don't have vehicles, then their ability to access playspaces is decreased. So areas that have higher percentage of households without vehicles should be somewhat prioritized for more playspaces.	American Community Survey	1.5
People in Households - without an Internet Subscription or no Computer per capita (2016-2020)	Proxy for income and access.	American Community Survey	1.5
Excessive Owner Housing Costs - 30 Percent or More of Income (2016-2020)	Proxy for livable / disposable income. Important complement to median family income since areas with high housing cost burden should be somewhat prioritized for more playspaces.	American Community Survey	1.5
HUD Subsidized Housing Units of properties (2021)	Proxy for income and access.	HUD Picture of Subsidized Household	1.5
Total number of child care providers (2022)	Proxy for higher demand for playspaces from ECE facilities.	PA Department of Human Services	1.5
Property With 2 or More Units of properties (2021)	Proxy for population density. More populous places receive higher priority.	American Community Survey	1
Pedestrian Road Network Density (2019)	Proxy for walkability (though a true measure of walkability for adults).	EPA Smart Location Database	1
Unemployment Rate (2016-2020)	Proxy for income. Weighted low because it is adult focused, and there may be challenges with its use during / post-pandemic with changes to labor force participation.	American Community Survey	1

People Commuting To Work - Public Transit per capita (2016-2020)	Proxy measure of access without personal vehicle. Those areas without as much access to public transit are prioritized.	American Community Survey	-1
Traffic Proximity and Volume (2020)	Proxy for safety and walkability. More walkable areas are prioritized.	EPA EJScreen: Environmental Justice Screening and Mapping Tool	-1

* Weights are either positive or negative depending on which values per variable are deemed to warrant more priority. For example, areas with both *higher* rates of children with disabilities and *lower* median household income are prioritized more to increase equity. So, the variable "share of children with disabilities" receives a positive weight while "median household income" receives a negative weight – in other words, areas with higher income receive less priority (and a lower score) than areas with lower income (which receive higher priority and higher scores).

** Child population dependent variable. Census tracts with low child populations (*sum_pop_u6*) with have more volatile prioritization results for these variables.

*** Custom variable calculated by the Child Care Playspace Accessibility Model (see below)

Child Care Playspace Accessibility Model

A particular aspect of playspace equity KABOOM! and CCAoA wanted to examine in the Philadelphia prioritization process was including a measure of how far away each child care facility in the city was to the nearest playspace. We defined "playspace" as *either an outdoor playground or play area(indoor or outdoor) with permanently installed equipment.*

First, in order to calculate these statistics, we needed comprehensive data on both child care facilities and playspaces. For the former, we downloaded licensed child care facilities from the opendataPA data catalog. For the latter, there was no single dataset that reflected the comprehensive, accurate distribution of playgrounds in Philadelphia. So, we created our own by merging data we obtained from Open Street Map (OSM), the Philadelphia Office of Children and Families (including playgrounds which are playful learning installations and that resulted from the Project Rebuild initiative), Philadelphia Parks and Recreation, the Philadelphia Housing Authority, and KABOOM!'s own past builds in the city (including some Play Everywhere project sites). The most compressive dataset was OSM thanks to KABOOM!'s digitization initiative in 2021 (see here and here). This dataset importantly contained all the playgrounds of Philadelphia Public Schools. The data downloaded from the geofabrik database also importantly represented the playgrounds as polygons which allowed us to calculate the "Percent Area Covered by Playgrounds" variable in the prioritization model.

All playground data from non-OSM sources were deduplicated and removed if OSM already adequately represented that particular playground. Remaining non-OSM playgrounds were buffered by a radius of 37 feet to create approximate polygons of the playground footprint. This radius was determined by a random sample of the playgrounds which these non-OSM points represented which had an average of 4,300 ft^2. The generated polygons were merged with the OSM polygons, resulting in a comprehensive playground footprint dataset for Philadelphia.

Philadelphia Playspace Mapping

Quantitative Methods Technical Brief

With both comprehensive provider and playground datasets for Philadelphia in-hand, we now were able to measure of how far away each child care facility in the city was to the nearest playspace. We used GIS to employ the Origin-Destination Cost Matrix Network Analysis to measure child care programs' enrolled children's walking distance to their nearest accessible playground. To construct the network dataset route layer in this analysis, we used OSM's walkable roads, paths, and sidewalks. To do this, we sorted through the complete list of values attached to the OSM key tag "highway" to determine which types of roads and paths were suitable for ECE teachers to walk on with their child care students. We selected the following value tags that fit this criteria:

'residential', 'secondary', 'secondary_link', 'tertiary', 'footway', 'tertiary_link', 'cycleway', 'pedestrian', 'track', 'steps', 'path', 'living_street', 'track_grade2', 'bridleway', 'track_grade5', 'track_grade1', 'unknown'

and excluded tunnels. We selected these walkable features from the geofabrik highway download data and created a network dataset with them.

The results were stored at the ECE provider level in miles. Originally, KABOOM! and CCAoA discussed employing a "reasonable walking distance" threshold to show providers that did and did not have a playground within a quarter-mile or 5 minute walk. However, to work fluidly with the other continuous variables used in the prioritization model, we left the results of the analysis as continuous values representing the total miles to the nearest playground.

Z-Score Distributions

We conducted the prioritization model at the smallest (most precise) unit of analysis possible : the census tract. Most demographic and socioeconomic variables existed at this level, but for those that were not, we calculated the average (playspace accessibility model) or raw (total percent area covered by playground, total number of child care providers) statistics for each tract city-wide.

Once every variable was calculated at the census tract level, we calculated the z-score for each to assess each tract's variation across the city and determine how unusual each one may be. In other words, we quantified how the census tracts' variable data values fell within percentiles compared to the average of all the city's tracts for those variable datasets. So, based on how spread out the data in the dataset are, z scores tell you relatively how exceptional each data point is. Using z scores was an ideal way for us to standardize into one unit of measurement a wide variety of source data across the 24 variables with unique units of measurements. Each z score calculation was conducted for each variable with a degrees-of-freedom value of n-1 with Null values excluded. True 0 values were retained where appropriate, i.e., where there truly was a value of nothing for a data point. Where measurements for data points simply did not exist, Null values were applied. 0 values were included and Null values were excluded from the z-score distribution calculations.

Prioritization Weights

Our final prioritization scores were calculated though an index of weighted z-scores. In order to prioritize variables that were more contextually important to making early learning equity decisions in Philadelphia, KABOOM! and CCAoA agreed upon and employed weights to emphasize some particularly critical data points. For example, KABOOM! holds that far too many kids lack adequate places to play due to the ongoing effects of systemic racism. Thus, we prioritize "minority population per capita" as a highly-weighted variable to help account for this. The following variables were weighted the highest:

- Minority population per capita
- Median household income
- Population ages 5 and younger per capita
- Existing playgrounds in the area, and
- The Child Care Playspace Accessibility Model's results

Weights were applied by multiplying each variable's z-scores by the designated weight. The absolute value of weights applied ranged from 1 to 2.5. Additionally, we applied some negative weights to standardize all variables with positive weighted z-scores representing the most important spaces and negative weighted scores representing the opposite. For example, to try to lift up children in low-income communities, we necessarily have to flip the weighted z-score of "median household income;" otherwise, communities with the *most* income would be prioritized.

Each weighted z score was summed to create one final prioritization score at the census tract unit of analysis. To help map users interpret these scores, we created an ordinal scale that ranges from "Lowest" to "Highest Priority Area." These categories were constructed as shown in Table 2:

Numeric Range	Ranges	Interpret	Frequency
-34.92 – -16.68	Minimum -P_{10}	Lowest Priority	40
-16.69 – -8.61	P_{10} – P_{25}	Low Priority	62
-8.62 – 0	P_{25} – 0	Somewhat Low Priority	85
0 – 8.2	0 – P_{75}	Somewhat High Priority	119
8.30 – 14.23	P_{75} – P_{90}	High Priority	61
14.24 – 41.92	P_{90} – Maximum	Highest Priority	41

You can view these prioritization scores, their categories, and their 24 disaggregated weighted z scores per census tract in our "Census Tract Priority Score" layer in the webmap.

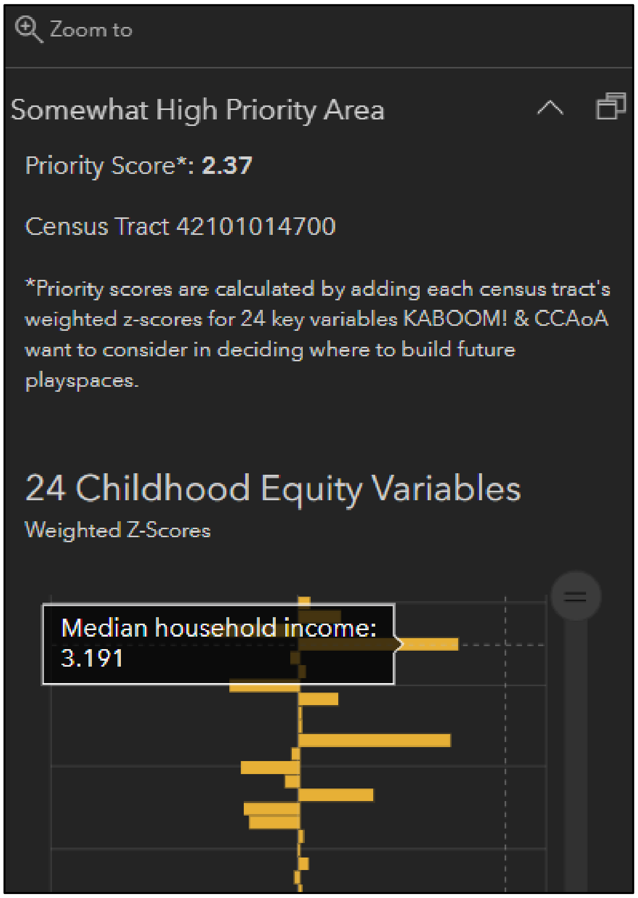

Figure 1: Example of a Census Tract's priority score and the variables' weighted z-scores contributing to it.

Upsampling to Neighborhoods and Planning Districts

Census tracts are good units of analysis to use in geographic data science because of the breadth of information that is tabulated at that scale and its precision in showing spatial variation of data across small areas of study (e.g., cities). However, tracts are not meaningful to the general population which can diminish their efficacy in translating the results of GIS data analyses to the public. In Philadelphia, neighborhoods are important local places that demarcate space and engender senses of belonging for residents. Additionally, city officials and planners use the city's official planning districts to make city-wide decisions and allocations of resources and funding.

In recognition of these realities, we sought to translate our census tract level results to these other more accessible geographies. We recognize that neighborhoods as vernacular regions are not spaces with fixed borders. A neighborhood is a more nebulous part of residents' mental map of Philadelphia that may have conflicting borders and senses of place depending on with whom you speak. To address this real world complication, we consulted with local officials and residents and will use the city's official neighborhood spatial dataset as of 2022 to demarcate neighborhoods for this project.

To move from census tract to neighborhood and planning district, we used the principles of the geographic analysis technique "upsampling" to scale-up the priority scores. Because tracts do not always fit neatly inside neither neighborhoods or planning districts, we could not perform simple calculations to move between scales. Instead, we converted the census tract's weighted priority scores to raster cells city-wide measuring 0.0011 decimal degrees[2]. For both neighborhoods and planning districts, we overlaid the borders on top of the priority score raster. We then calculated the average weighted priority value contained within each neighborhood and planning district to get an average priority score for each layer.

This way, each higher-level unit of analysis gets proportionally allocated the priority value from each of its intersecting census tracts. Thus, if a neighborhood was covered 75% by a tract with a priority score of 2.05 and another tract covered the remaining 25% with a priority score of -0.5, the neighborhood would have an average priority score of 1.4125 [(.75*2.05+.25*(-0.5))=1.4125]. While we're aware of the pitfalls of this methodology (re. Modifiable Areal Unit Problem), we feel it is a consistent and quantitatively sound way of transitioning scales with the substantial benefit of increasing public legibility of this work.

We conducted this upsampling technique with only one value: the final weighted priority score. While it would have been possible to upsample each contributing variable's weighted z-score, we chose to focus only on the final metric to direct attention to certain areas. At that point, census tracts contributing to neighborhood priority scores can be viewed for their variable-level scores.

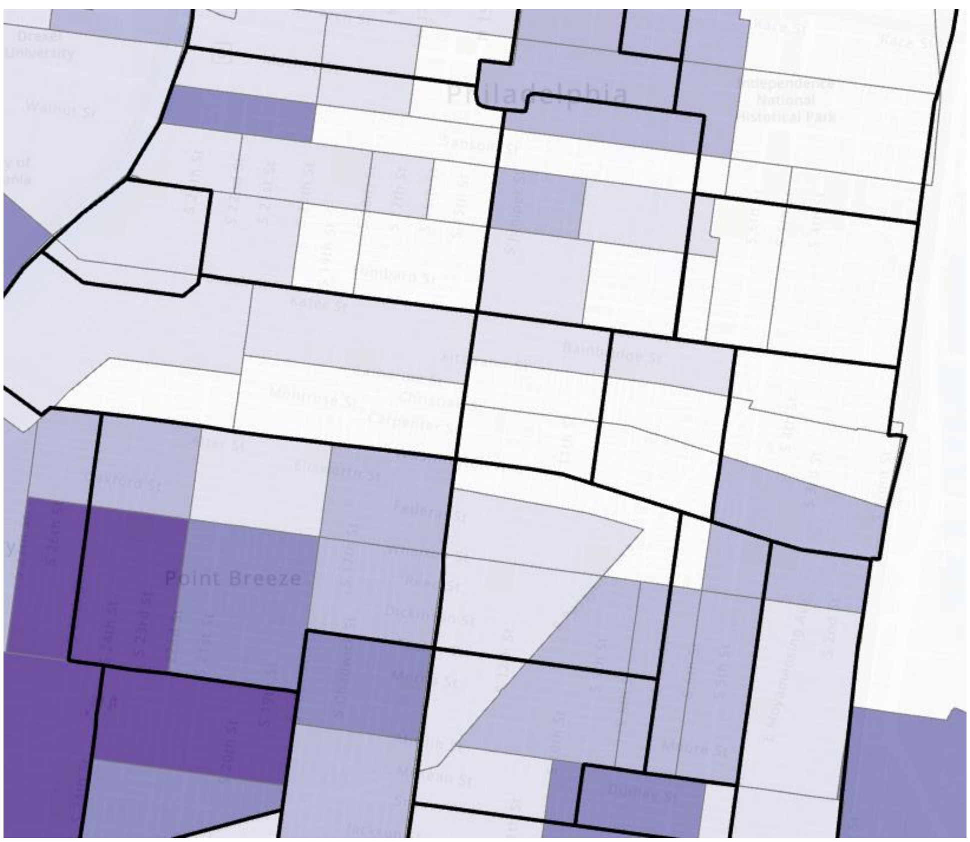

Figure 2: Neighborhoods (Black lines) do not match uniformly with Census Tracts (Purple-to-white polygons)

Philadelphia Playspace Mapping

Quantitative Methods Technical Brief

Data sources:

Below summarizes the data categories included on and excluded from the map.

Included

Dataset	Source(s)	Notes
Child care providers	OpenDataPhilly	
Potential child care demand	American Community Survey	Table B23008
Public Schools	OpenDataPhilly	
Playgrounds	OpenStreetMap (OSM), KABOOM!, Philadelphia Parks & Rec, Philadelphia Housing Authority	Resulted in a custom dataset of playground geographic areas. Metadata lost during merging process.
Socioeconomic Indicators	mySidewalk	Collation resource of different datasets including includes race, income, health, available park spaces, vehicle ownership, traffic volume. See each priority variable for original source.

Excluded

Dataset	Source(s)	Notes
Air Quality	OpenDataPhilly	Confusing dataset to interpret in the context of playspace equity. Poorer air quality spaces are often the result of environmental racism yet are less healthy play places.
Playspace quality	N/A	We do not have access to this key variable. While the Philadelphia Parks & Rec dataset sometimes had age of playspace attached, that would have only been a proxy of quality.
ECE Providers with Playgrounds	ECE Provider Survey	While our playground dataset should represent all outdoor playspaces in Philadelphia, including those of ECE centers, we were not able to definitively link playgrounds to providers using our ECE dataset and the results of our provider playspace survey.