Providence Census Geography Crosswalk

Brown University Community-Engaged Data and Evaluation Collaborative
Frank Donnelly, Head of GIS & Data Services, Brown University Library

Mar 18, 2025

https://github.com/Brown-University-Library/geodata_pvdcrosswalk

---

Introduction

The Providence Census Geography Crosswalk is a set of spreadsheet tables and a series of maps that relate census geographies with local geographies in Providence, Rhode Island. Census tracts, block groups, and ZIP Code Tabulation Areas (ZCTAs) for 2020 were related to Providence neighborhood boundaries and wards; the City's official neighborhood boundaries were regenerated using 2020 census blocks for this analysis. There is one crosswalk table for each pair, and each record represents an instance where a census geography and local geography overlap. The tables can be used for apportioning and aggregating census data from tracts and ZCTAs to neighborhoods and wards. A list of census blocks that constitute all of the geographies is also included. The crosswalk and maps, as well as the geodatabase used to generate them, are available on GitHub.

This product was produced as part of Brown University's Community-Engaged Data and Evaluation Collaborative (CEDEC), an initiative of the Swearer Center for Public Service that brings partners from the University and Rhode Island communities together to advance data-oriented projects. The Crosswalk was created by the GeoData@SciLi team, which is part of the Center for Library Exploration and Research (CLEAR) at the Brown University Library.

Disclaimer: Every effort was made to ensure that the data, which was compiled from public sources, was processed and presented accurately. The creators and Brown University disclaim any liability for errors, inaccuracies, or omissions that may be contained therein or for any damages that may arise from the foregoing. Users should independently verify the accuracy and fitness of the data for their purposes.

Table Structure

Columns in the crosswalk include values and percent totals for population, housing units, land area, and total area for the portion of the census geography that falls within the local geography. Population and housing counts are from the 2020 census, while area is from the 2023 Census TIGER Line files and is represented in square miles.

Records are sorted by census geography and then by local geography. Census tracts and block groups that are listed once in the table are fully contained within a neighborhood or ward; their percent total is equal to 1, meaning that 100% of its population, area, etc. is within that neighborhood or ward. Census geographies that appear more than once are split between multiple neighborhoods

or wards. The percent totals indicate the proportion of the census geography's population, area, etc. that falls within each neighborhood or ward. Values and percent totals for all split areas sum to 100%, as census tracts and block groups are fully contained within the City of Providence and do not cross City boundaries.

In the example below, census tract 1.01 is listed once in the tract2020_to_nbhood2020 crosswalk, which means it is fully contained within Washington Park. It's pct_pop value is 1, which means 100% of the population of tract 1.01 falls within Washington Park. Census tract 1.02 is listed twice, as it is split between South Elmwood and Washington Park. The total population of this tract is 5,527; 31.5% (1,741 people) live in South Elmwood, while 68.5% (3,786 people) live in Washington Park.

	A	B	C	D	E	F	G
1	**geoid_long**	**geoid_short**	**tractce20**	**tract_short**	**nbhood_2020**	**pop**	**pct_pop**
2	1400000US44007000101	44007000101	000101	1.01	Washington Park	4571	1
3	1400000US44007000102	44007000102	000102	1.02	South Elmwood	1741	0.315
4	1400000US44007000102	44007000102	000102	1.02	Washington Park	3786	0.685

The ZCTA tables are identical to the tract and block group tables, except that ZCTAs are not fully contained within the City's boundaries. Most of the values and percent totals for the ZCTAs do not sum to 100%, as these ZCTAs have population and area that fall outside the City. The correspondence between ZCTAs and local geographies is quite poor, and the tables and maps provided here illustrate this lack of correspondence. The tract and block group tables are better suited for generating census estimates.

In the example below, if we sum the pct_pop values for ZCTA 02905 in the zcta2020_to_nbhood2020 crosswalk, the value is .609, or 60.9%. This means that 39.1% of this ZCTA's population lives outside the City of Providence (population and area outside the City is not included in the table).

	A	B	C	D	E	F	G
1	**geoid_long**	**zcta_2020**	**nbhood_2020**	**pop**	**pct_pop**	**houses**	**pct_hous**
20	860Z200US02905	02905	Elmwood	48	0.002	16	0.002
21	860Z200US02905	02905	Lower South Providence	6556	0.249	2281	0.228
22	860Z200US02905	02905	South Elmwood	231	0.009	63	0.006
23	860Z200US02905	02905	Upper South Providence	838	0.032	387	0.039
24	860Z200US02905	02905	Washington Park	8357	0.317	2858	0.285

Crosswalk Tables and Columns

tract2020_to_nbhood2020: Relates 2020 census tracts to Providence neighborhood boundaries that were adjusted to coincide with 2020 census blocks.

tract2020_to_ward2022: Relates 2020 census tracts to Providence wards delineated in 2022.

bgroup2020_to_nbhood2020: Relates 2020 census block groups to Providence neighborhood boundaries that were adjusted to coincide with 2020 census blocks.

bgroup2020_to_ward2022: Relates 2020 census block groups to Providence wards delineated in 2022.

zcta2020_to_nbhood2020: Relates 2020 ZIP Code Tabulation Areas to Providence neighborhood boundaries that were adjusted to coincide with 2020 census blocks.

zcta2020_to_ward2022: Relates 2020 ZIP Code Tabulation Areas to Providence wards delineated in 2022.

blocks2020_pvd: List of 2020 census blocks in Providence with assignments to tracts, ZCTAs, neighborhoods, and wards

geoid_long	Long form of census geographic ID that contains summary level and FIPS codes
geoid_short	Short form of census geographic ID that contains FIPS code
tractce20	Six-digit 2020 census tract number
tract_short	Short form of tract number without trailing and leading zeros and with decimal
tract_bg	Seven-digit 2020 census tract and block group number
bgroup	One-digit census block group number
blockce20	Four-digit 2020 census block number
zcta2020	Five-digit 2020 ZIP Code Tabulation Area number
nbhood_2020	Providence neighborhood name
ward_2022	Providence ward number
pop	Population count from 2020 census
pct_pop	Percentage of total census population within Providence geography
houses	Housing unit count from 2020 census
pct_hous	Percentage of total census housing units within Providence geography
land_sqmi	Land area in square miles from 2023 Census TIGER Line Files
pct_land	Percentage of census geography land area within Providence geography
area_sqmi	Total area in square miles from 2023 Census TIGER Line Files
pct_area	Percentage of census geography area within Providence geography

Using the Crosswalk

The crosswalk allows you to take census data published at the tract, block group, and ZCTA levels (the ZCTA level is not recommended), and reformulate it to create summaries at the neighborhood and ward levels for the City of Providence:

1. Take a tract, block group, or ZCTA-level data table from the Census Bureau's decennial census or American Community Survey (ACS) and join it to the crosswalk table using the unique ID that the tables share in common. This will be a one-to-many join, where one record in the data table is related to one or many records in the crosswalk.

2. Choose a variable for apportioning the data (population, housing, land area, or total area), and multiply the census variables by the percent total for that apportionment variable.

3. Aggregate or group the results by neighborhood or ward to create the estimate for the local geography.

These steps can be achieved in a spreadsheet program like Excel by using VLOOKUP for relating the tables, sheet formulas for multiplication, and a Pivot table for grouping. In a relational database like SQLite, these steps can be achieved by using JOIN, creating calculated fields, and applying an aggregate SUM function with GROUP BY. Statistical packages like STATA and scripting languages like R and Python have similar functions for joining, creating variables, and aggregating by categories.

Caveats in using the crosswalk:

1. Derived statistics that represent means, medians, percent totals, or ratios cannot be apportioned in this manner. In these cases it's necessary to get the original values and calculate the derivatives after aggregating them. For example, to generate estimates for mean household income, you would need to apportion data for total number of households and total aggregate household income, aggregate those values to neighborhoods or wards, and then divide total income by households to generate the new mean household income values.

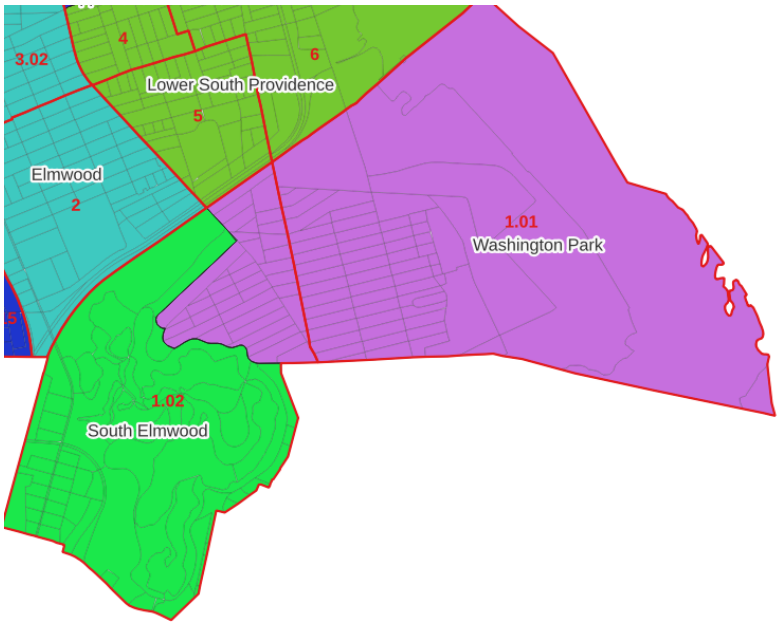
2. Estimates from the ACS are accompanied by a margin of error that indicates the possible range of the estimate. When creating new estimates, it's a best practice to generate new margins of error for those estimates so that the precision and quality of the estimate can be evaluated. Consult the Census Bureau's ACS Handbooks or their educational material at the Census Academy for details.

Methodology

2020 block, block group, tract, ZCTA, and county subdivision boundaries were sourced from the US Census Bureau 2023 TIGER/Line Shapefiles. The attribute tables included land and water area in square meters, which we converted to square miles. The block file included 2020 census counts of population and housing units. The Providence GIS Hub was used to obtain boundary files for 2022 wards and neighborhoods. The neighborhood file had a date stamp from 2013.

The crosswalk was generated from census blocks, which are the smallest unit of census geography that constitute the "building blocks" for all census geographies. Blocks were assigned the tract, block group, and ZCTA identifiers that they are part of. The local geographies for the City of Providence were overlaid onto the blocks using QGIS, and each block was assigned to the neighborhood and ward that contains it. A SQLite / Spatialite relational database was used to sum the blocks into distinct census geography / local geography combinations, and to calculate the percentage of the census geography's attributes that fell inside each local geography.

In the example below, the individual neighborhoods have distinct colors. The small areas symbolized with thin grey lines are census blocks, and the larger areas outlined in red with numbers are census tracts. Tract 1.01 falls entirely within Washington Park, while tract 1.02 is split between Washington Park and South Elmwood. Knowing the tract and neighborhood that each block falls within, we can aggregate the blocks for those distinct areas to get the values and percent totals for tract population, area, etc. within each neighborhood.



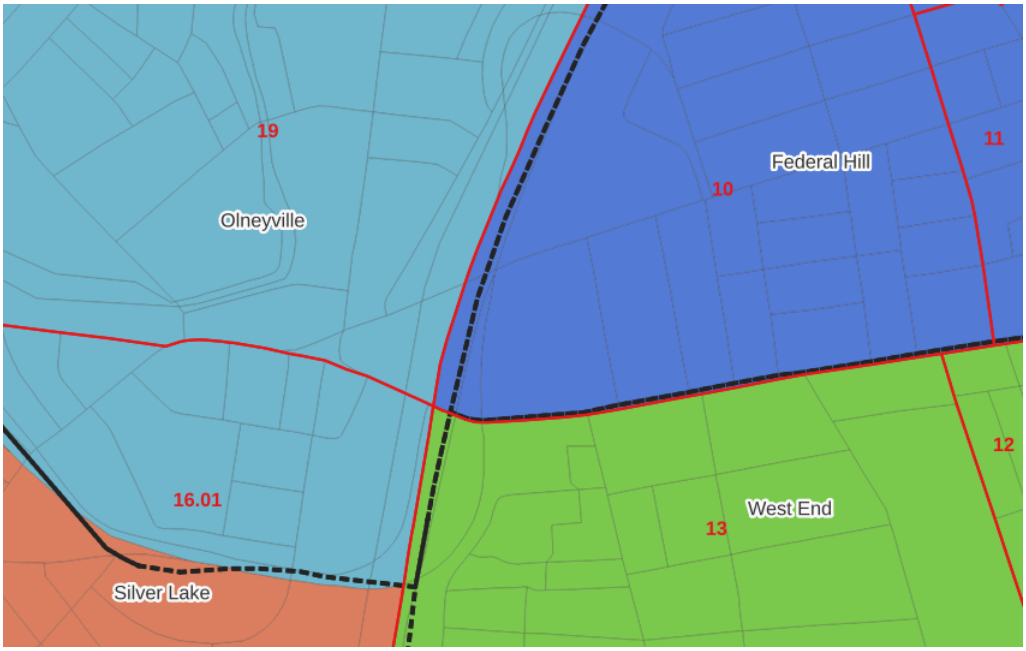
The City constructed the 2022 wards by aggregating 2020 census blocks, so the relationship between the census block files and the wards was identical, in that the blocks fit perfectly within ward boundaries without overlap.

Conversely, the neighborhood boundaries were constructed either from older census geography or from boundaries delineated by streets, and the blocks did not nest perfectly within the neighborhoods. The lack of correspondence was primarily due to changes in how unpopulated blocks were delineated over time, as well as to improvements in accuracy in drawing the 2020 block boundaries.

To account for these discrepancies, we constructed a new set of neighborhood boundaries using the 2020 census blocks. Blocks were initially assigned to the neighborhood they primarily fell within. Blocks along the borders of neighborhoods were subsequently scrutinized and re-assigned to avoid splitting non-populated blocks that were part of the same census tract into different neighborhoods. In almost all instances, these unpopulated areas consisted of highways, water bodies, and greenspace.[1] Non-populated blocks were not reassigned if they were sufficiently large enough to drastically alter the boundary of the neighborhood. A balance was struck to keep the new boundaries as close to the originals as possible, while correcting for the inaccuracies of non-matching block and neighborhood boundaries. These 2020 neighborhood boundaries were used for constructing the crosswalk.

In the example below, the black dotted line represents the original neighborhood boundaries. The 2020 boundaries for Federal Hill and the West End were moved slightly to the west of the original, to avoid assigning the long, thin, unpopulated blocks (covering a highway) that are part of tracts 10 and 13 to neighboring Olneyville and Silver Lake, which contain no other blocks from those tracts. The irregular boundary between Olneyville and Silver Lake is an example of the inconsistency between the delineation of the 2020 blocks with whatever boundaries were used in creating the original neighborhoods.

[1]The only exceptions were two populated blocks that were split by neighborhood boundaries: block 1006 in tract 16.01 (population 87) was assigned to Silver Lake instead of Hartford, and block 3011 in tract 1.02 (population 77) was assigned to Washington Park instead of South Elmwood. The reassignment was based on satellite imagery that showed that all the housing units in these blocks fell within the reassigned neighborhoods.



Given the reassignment of these unpopulated blocks, the land and total area of the 2020 neighborhoods will differ from the original, official neighborhoods delineated by the City of Providence. However, this reassignment has no effect in estimating the population and housing counts for neighborhoods, because the blocks are unpopulated (and for the few exceptions of split populated blocks, the population resided entirely within one neighborhood).