dataworks nc

DURHAM NEIGHBORHOOD COMPASS
DATA DOWNLOAD

The **Neighborhood Compass** at https://compass.durhamnc.gov is a publicly-accessible neighborhood data portal for residents of Durham city and county. It serves more than 60 variables related to community health, housing, transportation and demographics. It is maintained and developed by DataWorks NC and hosted by the City of Durham.

Data for the Compass comes from the United States Census Bureau, the City and County of Durham, Duke Health, Lincoln Community Health Center and DataWorks' own calculations.

If you have comments or questions about the download or how to use the data please don't hesitate to contact us at tech@dataworks-nc.org!

Contents of the Download

GEOGRAPHY FILES

The **geography** directory contains the following subdirectories:

blockgroup - Blockgroup geography shapefiles as found in the Census TIGER dataset.

tract - Census tract geography shapefiles as found in the CENSUS TIGER dataset

These geography files are also available in GeoJSON format at the following URLs:

Blockgroup - https://compass.durhamnc.gov/data/blockgroup.geojson.json

Tract - https://compass.durhamnc.gov/data/tract.geojson.json

DATA DICTIONARY - **DurhamNeighborhoodCompass-DataDictionary.csv** contains a list of all the metrics included in the data download, as well as links to more detailed metadata for each of them. It has the following fields

Metric title	Title of the metric
Metric category	Broad category (used in the Compass menus)
Metric code	Short code used to identify the metric within the Compass
Aggregation type	How this metric should be aggregated across multiple geographies. This can have one of 3 values:

	• **weighted** – Aggregate using a weighted average, using the values in the weights file (see below for more info) • **mean** – Just take a normal average of the values for each geography to get a combined value • **sum** – Add up the values of each geography to get a combined value
Label for the metric value	A descriptive label which can give more context to the metric value
Label for the raw value	For some **weighted** metrics which are percentages, multiplying the metric value by the value in the weights file gives a "raw" value which counts the total # of cases, incidents, people, etc. This label helps interpret what the raw value means
Geographies available	List of geographic levels for which the compass has data for this metric
Metadata link	Link to more information about this metric

DATA FILES - **DurhamNeighborhoodCompass-tract.csv** and **DurhamNeighborhoodCompass-blockgroup.csv**:

These data files, in CSV format, contain all the indicator values for each year. There's one row for each different geography (each tract or each blockgroup), and each column contains the values for a single indicator for a single year.

Geographies are keyed by their US Census FIPS codes in the "Geography ID" column, but we also include a more human-readable "Geography Label" column.

WEIGHTS FILES - **DurhamNeighborhoodCompass-tract-weights.csv** and **DurhamNeighborhoodCompass-blockgroup-weights.csv**

These files have the same format as the data files in terms of rows and columns, but instead of the data values themselves the cells contain weight values for the corresponding indicator/year/geography. These weight values are intended to help with doing statistical analysis of the indicator values, and especially in aggregating or grouping values across different geographic units. See below for more information.

All variables that are normalized by area will have a decimal area value for each year in the weights file. This is the area of the blockgroup or tract in square miles. All variables which are percentages of the population will have an integer value for each year in the weights file. This is the total population of the blockgroup for that year.

What is a weight value?

Many of the indicators in the Compass (and the data files) are expressed as ratios or percentages -- for example, the percentage of residents who were counted as African-American, the percentage of households within walking distance of a grocery store, or the number of residential occupancy permits per square mile.

In order to find the average value of an indicator across multiple areas, *we don't recommend averaging the indicator values directly*. Instead, you should create a *weighted* average by multiplying each indicator by the weight given in the weights file, adding up the different values, and then dividing the total by the sum of the weights.

Say you were researching the % youth population in the Lakewood area, and trying to combine Block Groups 2 and 3 in Tract 6.00. Here's the data that you'd find in the blockgroup data files for those two areas:

	Youth Population, 2010 (in DurhamNeighborhoodCompass-blockgroup.csv)	Weight for Youth Population, 2010 (in DurhamNeighborhoodCompass-blockgroup-weights.csv)
Block Group 2	25.63%	2,228 *<= this is the total # of people in the blockgroup in 2010*
Block Group 3	19.11%	518 *<= this is the total # of people in the blockgroup in 2010*

If you just average those two numbers, you'd get an average value of 22.37% young people over the whole area. But that gives unfair weight to Block Group 3, which has 1/4 of the population of Block Group 2. The correct way to combine these two values would be to use the weights as follows:

$$Youth\ population = \frac{BG\ 2\ youth\ pop * BG\ 2\ weight + BG\ 3\ youth\ pop * BG\ 3\ weight}{BG\ 2\ weight + BG\ 3\ weight}$$

$$Youth\ population = \frac{25.63\%\ youth * 2{,}228\ people + 19.11\%\ youth * 518\ people}{2228\ people + 518\ people}$$

$$Youth\ population = \frac{670\ young\ people\ total}{2746\ people\ total} = \mathbf{24.39}\%$$

Notice how the resulting value there is larger than what we got from just taking the average of the two percentages? Using a weighted average helped us account for the fact that the two different blockgroups had different populations to get an accurate total number of young people in the combined area.