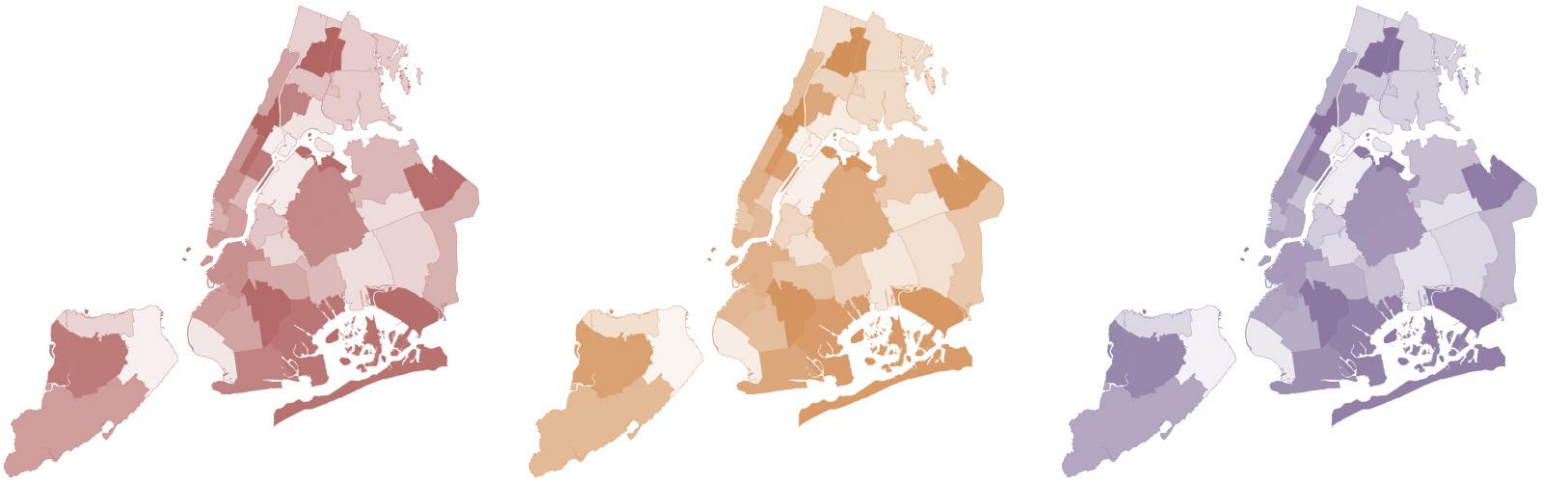
GIS Code Book: New York City United Hospital Fund (UHF) 42 Metrics



♠ Last Updated: 12-December-2014

Prepared for:

New York City Department of Health and Mental Hygiene (DOHMH)

Prepared by:

James W. Quinn, MA
Geographer, Senior GIS Analyst
Built Environment and Health Working Group
Department of Epidemiology
Mailman School of Public Health
Columbia University
e: jq2145@columbia.edu | p: (218) 343-6371
www.beh.columbia.edu

Columbia University
Built Environment & Health Project
www.beh.columbia.edu

United Hospital Fund (UHF) 42 Variables

uhf_id – Unique ID assigned by James Quinn for GIS processing.
uhf_code – Unique 3-digit United Hospital Fund 42 (UHF) ID.
uhf_name – UHF42 descriptive name.
uhf_km2 – Total area of UHF42 geography in km2.
uhf_lndkm2 – Total **land** area of UHF42 geography in km2 *(inland water bodies removed).*

uhf_code	uhf_name
0	Parks, Airports, Islands
101	Kingsbridge - Riverdale
102	Northeast Bronx
103	Fordham - Bronx Park
104	Pelham - Throgs Neck
105	Crotona - Tremont
106	High Bridge - Morrisania
107	Hunts Point - Mott Haven
201	Greenpoint
202	Downtown - Heights - Slope
203	Bedford Stuyvesant - Crown Heights
204	East New York
205	Sunset Park
206	Borough Park
207	East Flatbush - Flatbush
208	Canarsie - Flatlands
209	Bensonhurst - Bay Ridge
210	Coney Island - Sheepshead Bay
211	Williamsburg - Bushwick
301	Washington Heights - Inwood
302	Central Harlem - Morningside Heights
303	East Harlem
304	Upper West Side
305	Upper East Side
306	Chelsea - Clinton
307	Gramercy Park - Murray Hill
308	Greenwich Village - Soho
309	Union Square - Lower East Side
310	Lower Manhattan
401	Long Island City - Astoria
402	West Queens
403	Flushing - Clearview
404	Bayside - Little Neck
405	Ridgewood - Forest Hills
406	Fresh Meadows
407	Southwest Queens
408	Jamaica
409	Southeast Queens
410	Rockaway
501	Port Richmond
502	Stapleton - St. George
503	Willowbrook
504	South Beach - Tottenville

Built Environment Variables

Street Pattern Variables

_strint** – Count of "unique" street intersections excluding intersections with a valence of 1 or 2 (e.g., a valence of 1 indicates a dead-end or cul-de-sac; a valence of 3 indicates a three-way intersection, etc.).

_intden** – Density of "unique" streets intersections in km2 [**_strint / **_lndkm2].

Public Transportation Variables

Subway Stations

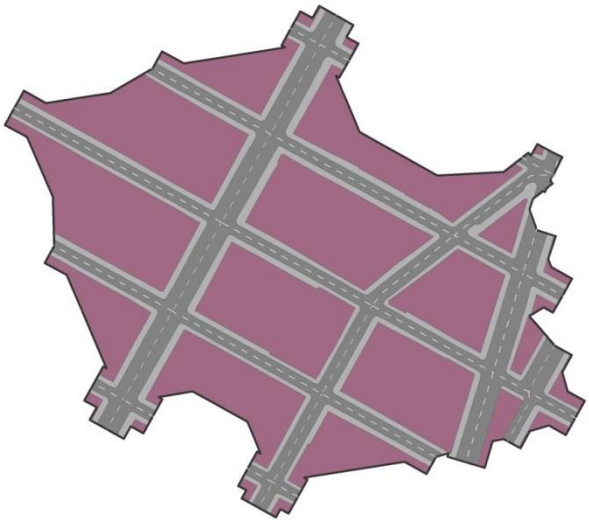
♠ The New York City **Subway** information was originally obtained from the New York City Transit Authority in 1998. The data was realigned to the NYCMap basemap in 2002. In November of 2006, a major shape update included PDF Subway System maps being downloaded from the MTA/NYC Transit web-site. Community Cartography updated the existing GIS files using on-screen digitizing and the downloaded subway maps. In August of 2007, station locations were spatially adjusted to align to the NYCMap streets at a map scale of 1:2,400. For more information, please visit the MTA Transit web-site at: www.mta.nyc.ny.us/nyct/maps/subwaymap.pdf.

_sub12c** – Count of MTA subway stations as of 2012. This measure takes into account those subway stations that have multiple route-transfer opportunities at one location and each stop is counted only **once** regardless of how many route-transfer opportunities are available at any given subway station.

_sub12d** – Density of subway stations per km2 [**_sub12c /**_lndkm2].

Roadbed & Sidewalk Variables

_strkm2** – Total roadbed area (i.e., curb-to-curb) in km2 within the neighborhood buffer definition (dark gray area →)
_strpct** – Percent of neighborhood buffer definition covered by roadbed area. [**_strkm2 / **_lndkm2].



_wlkkm2** – Total sidewalk area (i.e., curb-to-building frontage) in km2 within the neighborhood buffer definition (light gray area →).
_wlkpct** – Percent of neighborhood buffer definition covered by sidewalk area. [**_wlkkm2 / **_lndkm2].

Tree Canopy Variables

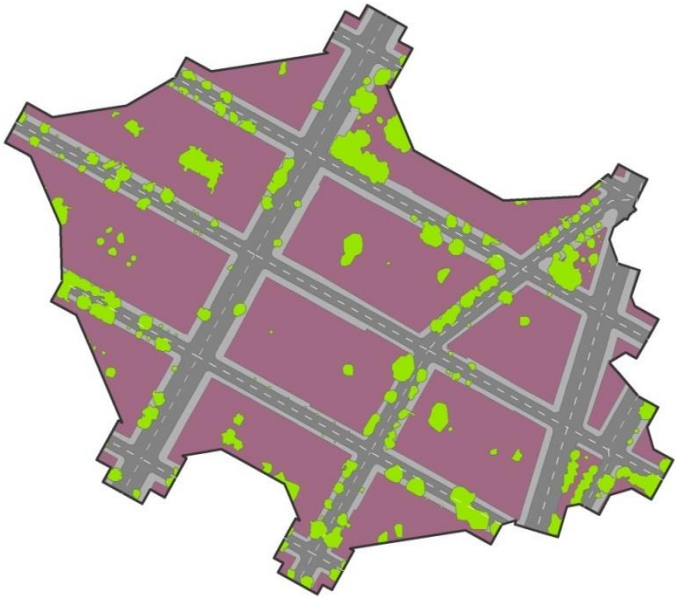
The 'Tree Canopy' dataset used for this project was developed as part of the Urban Tree Canopy (UTC) Assessment for NYC. As such, it represents a 'top down' mapping perspective in which tree canopy over hanging other features was assigned to the tree canopy class. The minimum mapping unit for the delineation of features was set at 3-ft2. The primary sources used to derive this dataset were 2010 LiDAR and 2008 4-band orthoimagery data. Ancillary GIS data sources included city boundaries, building footprints, water, parking lots, roads, railroads, railroad structures, and ball fields. This land cover dataset is considered current as of 2008.

For each neighborhood buffer definition, the amount of tree canopy was quantified three ways: **1**) amount within the 'entire buffer' area; **2**) amount within the 'roadbed' area within the buffer; and **3**) amount within the 'sidewalk' area within the buffer.

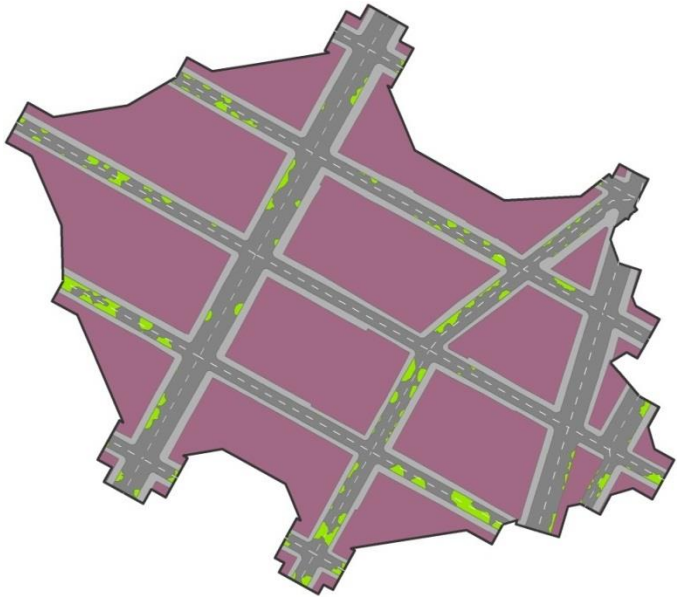
_cpykm2** – Area of tree canopy in km2 within the entire neighborhood buffer definition.
_cpypct** – Percent of entire neighborhood buffer definition covered by tree canopy [**_cpykm2 / **_lndkm2].

_rbckm2** – Area of tree canopy in km2 within the 'roadbed' area of the neighborhood definition.
_rbcpct** – Percent of neighborhood definition covered by tree canopy [**_strkm2 / **_lndkm2].

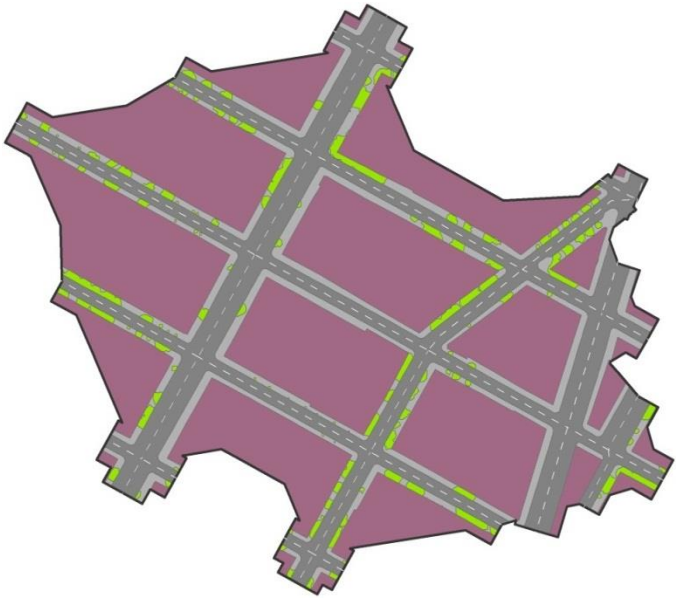
_swckm2** – Area of tree canopy in km2 within the 'sidewalk' area of the neighborhood definition.
_swcpct** – Percent of neighborhood definition covered by tree canopy [**_wlkkm2 / **_lndkm2].



method 1: 'entire buffer' area



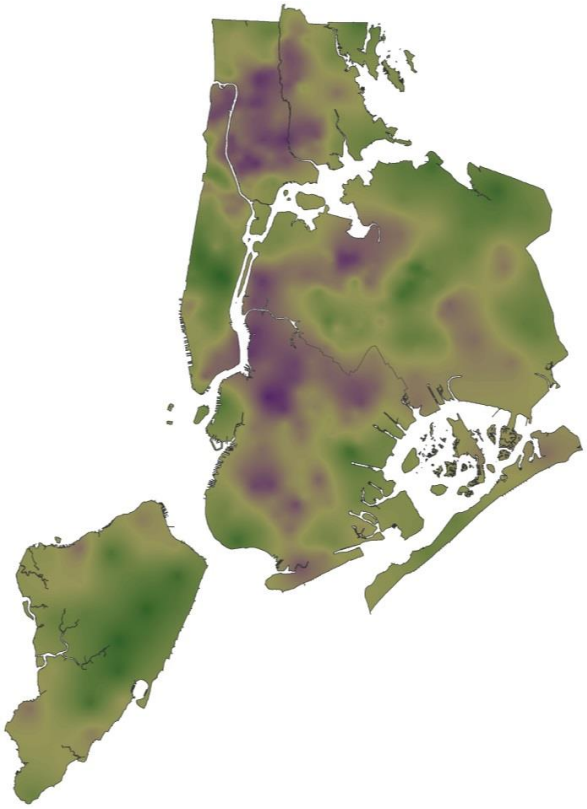
method 2: 'roadbed' area



method 3: 'sidewalk' area

Neighborhood Physical Disorder Variables

Neighborhood physical disorder, or the deterioration of an urban landscape, is reflected in indicators such as litter, graffiti, and abandoned buildings. A physical disorder surface was constructed from observations of nine indicators of disorder (i.e., bottles, litter, graffiti, poorly maintained buildings, burned out buildings, abandoned buildings, abandoned cars, vacant lots, and bars on windows) on 532 block faces in NYC using Google Street View imagery that dated from August 2007 to October 2011. The nine indicators were combined into a scale using Item Response Theory, and ordinary kriging with an exponential variogram function with nugget = 0.19, range = 5000 and partial sill = 0.14 was used to estimate physical disorder levels at any point throughout the city.



For each UHF42 geography, 11 neighborhood disorder variables were constructed; one from each of 10 kriged conditional realizations and one from the kriged point estimate ('value'). Because kriging presents over-smoothed estimates, when estimating associations with disorder, rather than use the point estimate, the 10 conditional realizations should be used as imputations within a multiple imputation framework. This will mitigate the over-smoothing issue and result in associational estimates that account appropriately for uncertainty and measurement error both in the point disorder measure and the interpolation. The simulations are conditional realizations based on the kriging parameters, observed values, and variance estimates – essentially, the conditional realizations use estimates that sample from the posterior distribution rather than select the peak of that distribution.

_svd_c** – Count of 50-by-50-foot kriged grid cells falling within the spatial bounds of each UHF neighborhood definition.
_svd_m** – Mean kriged point estimate ('value').
_sim1_m** – Mean kriged conditional realization; simulation 1.
_sim2_m** – Mean kriged conditional realization; simulation 2.
_sim3_m** – Mean kriged conditional realization; simulation 3.
_sim4_m** – Mean kriged conditional realization; simulation 4.
_sim5_m** – Mean kriged conditional realization; simulation 5.
_sim6_m** – Mean kriged conditional realization; simulation 6.
_sim7_m** – Mean kriged conditional realization; simulation 7.
_sim8_m** – Mean kriged conditional realization; simulation 8.
_sim9_m** – Mean kriged conditional realization; simulation 9.
_sim10_m** – Mean kriged conditional realization; simulation 10.

no signs of disorder

empty bottles

graffiti

bars on windows

vacant land

adondaned cars

litter

poorly maintained buildings

burned out buildings

abandoned buildings

Land-Use Mix Variables

Land-use mix is a measure of different activities or different destinations. Mixed land-use is important as it can provide a greater variety of attractive destinations within a walking distance and more visual variety and interest for pedestrians as varied land-uses are viewed as promoting architectural and landscape variety. Mixed land-use can also be associated with greater street safety due to informal policing—one is less likely to be alone with an attacker. This assumes that uses have a mixture of opening hours, particularly in the evening, and generate pedestrian traffic. Some uses may be perceived, however, to undermine safety—e.g., rowdy bars or vacant land/buildings.

The land-use mix variables calculated for this project look at total residential and commercial building areas, a co distribution of commercial and residential building area, and a simple measure of land-use percentage in 11 major uses assigned by The NYC Department of City Planning at the parcel level, which is less a measure of mix per se than the relative amount of each land-use:

- **01**. One & Two Family Buildings;
- **02**. Multi-Family Walk-up Buildings;
- **03**. Multi-Family Elevator Buildings;
- **04**. Mixed Residential and Commercial Buildings;
- **05**. Commercial and Office Buildings;
- **06**. Industrial and Manufacturing;
- **07**. Transportation and Utility;
- **08**. Public Facilities and Institutions;
- **09**. Open Space and Outdoor Recreation;
- **10**. Parking Facilities; and
- **11**. Vacant Land.

_reskm2** – Residential building area in km2 (**_lumixa input variable).
_respct** – Percent of residential building area [([**_reskm2] / ([**_comkm2] + [**_reskm2])) * 100].
_comkm2** – Commercial building area in km2 (**_lumixa input variable).
_compct** – Percent of commercial building area [([**_comkm2] / ([**_comkm2] + [**_reskm2])) * 100].
_lumixa** – Land Use Mix A. This measure was calculated using the MapPLUTO (version 10v2; March 2010 – September 2010) tax lot data available from the NYC Department of City Planning. A co-distribution of commercial and residential building area was derived from the MapPLUTO data as an indicator of neighborhood walkability. Building area was used rather than land area was used because in dense, mixed-use environments it is often impossible to designate a building as entirely one land use or another. In lower density areas where buildings are more likely to be single-use structures, building area will be equivalent to land area. A simple index was constructed varying between zero and one that captures this relationship. Building areas in each category are summed up to the measurement geography of analysis and divided by the total of the two building areas. These two ratios are then multiplied by one another, and then scaled by a factor of four so that the range of the index will go between zero and one. In a perfectly mixed area – containing equal areas of residential and commercial space – this index is equal to one. If either area dominates, the index will tend towards zero. The following equation describes this relationship.

$$LM_A = 4 \cdot \left(\frac{\sum A_{res} \cdot \sum A_{comm}}{\left(\sum \left(A_{res} + A_{comm}\right)\right)^2} \right)$$

Walkability Index Scale

A number of researchers have constructed walkability indices which summarize built environment features believed to promote walking. Although specification details vary, these indices typically include measures of population density, land use, and street network. Our walkability measure was adapted from that employed in recent papers by Frank and colleagues (2005 and 2006), which includes four components: residential population density (density of population per total residential land area), intersection density, an entropy measure of land use based on the distribution of building floor area among six land use types (education, entertainment, single-family residential, multi-family residential, retail, and office), and the retail floor area ratio, or the ratio of retail building floor area to retail land area. All of the Frank components were z-scored and summed, with intersection density receiving a double weight for the *Frank Scale*, but not for our scale. Our "BEH walkability scale" is documented in a paper by Neckerman and colleagues (2009).

Frank et al. Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. American Journal of Preventive Medicine. 2005;28(2 Suppl 2):117-125.

Frank et al. Many Pathways from Land Use to Health: Associations between Neighborhood Walkability and Active Transportation, Body Mass Index, and Air Quality. Journal of the American Planning Association. 2006;72(1):75-87.

Neckerman et al. Disparities in urban neighborhood conditions: Evidence from GIS measures and field observation in New York City. Journal of Public Health Policy. 2009;1(1 Suppl): S264-S285.

Walkability Index Scale "Component" Variables

_intden** – Density of unique streets intersections per km2.
_sub12d** – Density of subway stations per km2.
_rtlfar** – Retail floor area ratio – Retail building floor area divided by retail land area in km2.
_resdn1** – Density of res units – Number of residential units divided by total residential building floor area in km2.
_entrpy** – Land Use Mix – An entropy measure using the five of the six land use types employed in Frank et al. (2006). Single- and multi-family residential areas were combined because most housing in New York City is multi-family. Parcel-level measures of residential, office, and retail floor area were available from the MapPLUTO (version 10v2; March 2010 – September 2010) database. We used the MapPLUTO building class codes to identify buildings associated with education (schools) or entertainment (theaters, recreational facilities), and attributed the entire floor area of the identified building to education or entertainment. The entropy formula used was adapted from Frank et al. (2005), which yielded more plausible results: Land Use Mix = A/ln(N) where: A= –((b1/a)*ln(b1/a)+(b2/a)*ln(b2/a)+…) and b1 is the building floor area covered by the first land use, b2 is the building floor area covered by the second land use, etc., a is the total floor area across the five land uses, and N is the total number of land uses considered (i.e., 5) in the UHF. Zero values for b1…b5 were set to .000001 to avoid zero or undefined terms.

Components of the Entropy Measure*:*

_b1 –** Total building area for Education uses in square feet (set to .000001 if 0)
_b2 –** Total building area for Entertainment uses in square feet (set to .000001 if 0)
_b3 –** Total building area Residential uses in square feet (set to .000001 if 0)
_b4 –** Total building area for Retail uses in square feet (set to .000001 if 0)
_b5 –** Total building area for Office uses in square feet (set to .000001 if 0)
_a –** Total floor area across the five land uses in square feet (set to .000001 if 0)
_n –** The total number of land uses considered (i.e., 5)

ArcMap Entropy Field Calculation Expression:

entropy = –((([b1] / [a]) * log ([b1] / [a])) + (([b2] / [a]) * log ([b2] / [a])) + (([b3] / [a]) * log ([b3] / [a])) + (([b4] / [a]) * log ([b4] / [a])) + (([b5] / [a]) * log ([b5] / [a]))) / log ([n])

Walkability Index Scale Variables

To date, BEH has created and used two different versions of the Walkability Index Scales, which we refer to here as the *"Frank et al. 2006"* and BEH scales. The *"Frank 2006"* includes z-scored variables: residential density, land use mix using 5 land use types, intersection density * 2, and retail area ratio. The BEH scale includes z-scored variables: residential density, land use mix using 5 land use types, intersection density, retail area ratio, and subway stop density. Note that the BEH scale does not multiply intersection density by 2 and does include subway density.

****_intden_z** – Density of unique streets intersections per km2: z-scored.
****_sub12d_z** – Density of subway stations per km2: z-scored.
****_rtlfar_z** – Retail floor area ratio: z-scored.
****_resdn1_z** – Density of residential units: z-scored.
****_entrpy_z** – Entropy land use mix: z-scored.

****_beh_walk** – BEH Walkability Scale.
****_beh_walk_cat** – Quintiles of BEH Walkability Scale.

Walkability Index Scale by UHF42