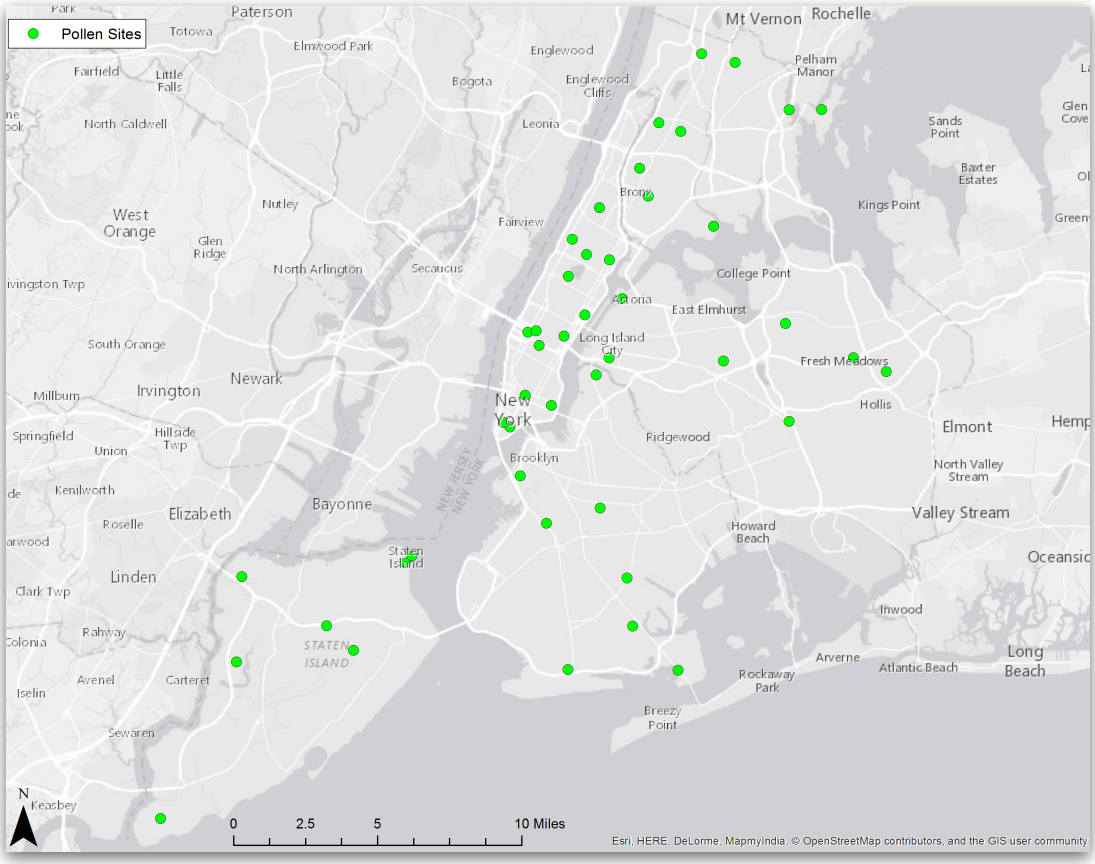
Tree Pollen Sites and Grid points (100 meter) Variables Data Dictionary for NAAS



Prepared By:

Daniel M. Sheehan
Geographer, GIS Analyst
Built Environment and Health
Department of Epidemiology
Columbia University
722 West 168th Street, R735
New York, New York 10032
www.beh.columbia.edu
dms2203@cumc.columbia.edu

Prepared For:

Kate Weinberger
PhD Candidate, Department of Environmental Health Sciences, Columbia University
New York, New York 10032
www.beh.columbia.edu
krw2114@columbia.edu

Prepared For:

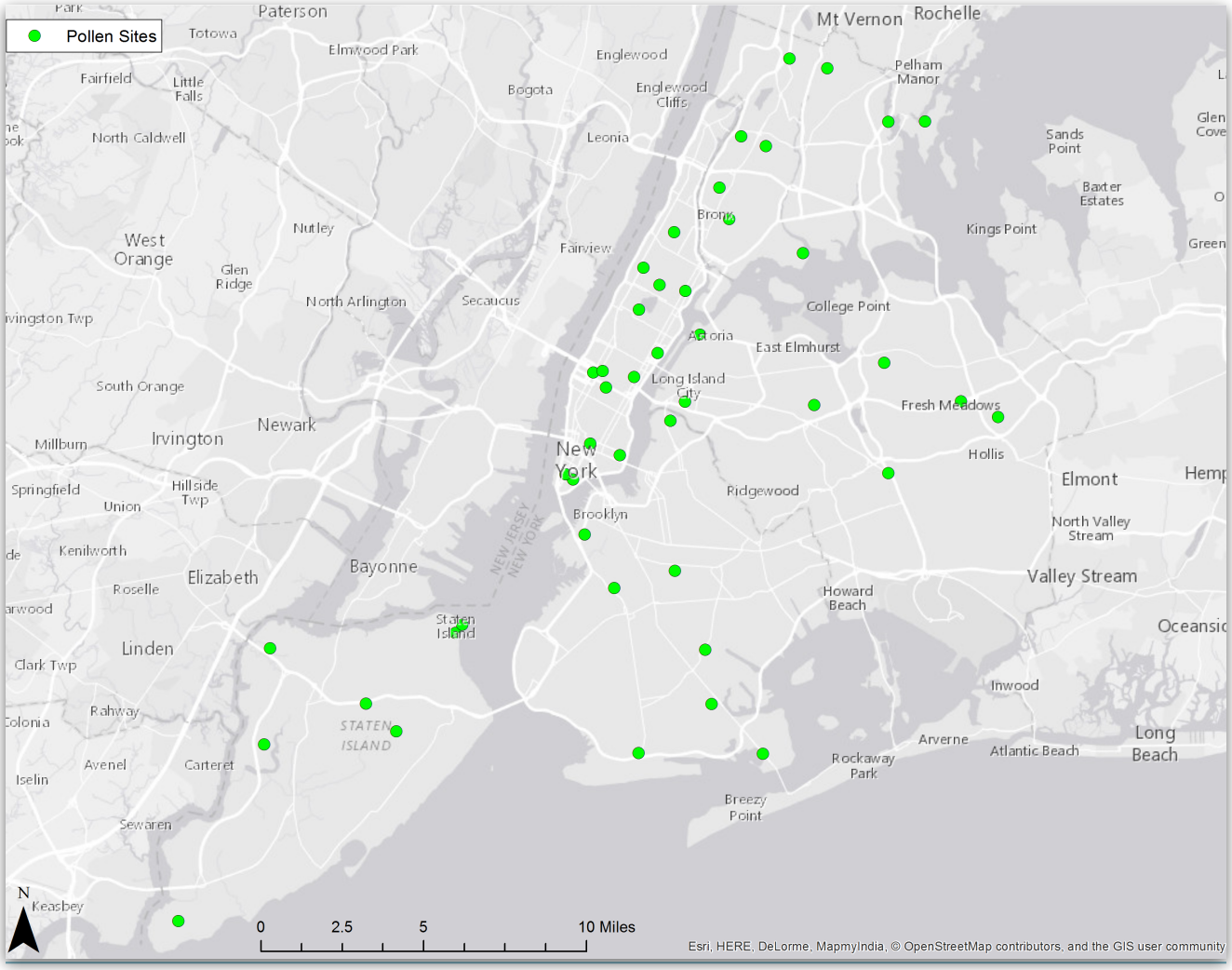
Dr. Gina Lovasi
Assistant Professor of Epidemiology
Built Environment and Health
Department of Epidemiology
Columbia University
New York, New York 10032
www.beh.columbia.edu
glovasi@columbia.edu

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

ABOUT

The following document is the data dictionary for variables pertaining to 45 locations that were used for pollen monitoring. Kate Weinberger (krw2114@columbia.edu), PhD Candidate, at the Department of Environmental Health Sciences at the Mailman School of Public Health at Columbia University administered the site selection and pollen sampling. This data dictionary's purpose is to aid in her analysis of the pollen sites' and grid points data. These pollen sites and future analysis are being performed along with the Neighborhood Asthma and Allergy Study (NYC NAAS) study.

Map of Pollen Sites



DATASETS

There are unique datasets for the following point geographies:

Point Geographies

Pollen Sites

There are **45** pollen monitoring sites that were set up by Kate Weinberger, PhD Candidate, at the Department of Environmental Health Sciences at the Mailman School of Public Health at Columbia University. These sites had a latitude and longitude and were then projected into the project's primary coordinate system, NAD 1983 StatePlane New York Long Island FIPS 3104 Feet.

Filename: **pollen_sites.csv**

Grid Points

The grid points dataset originates from the NYCCAS Air Pollution data. NYCCAS sent a Shapefile called NYCCAS_100m_grid.shp. This file was a vector (Shapefile) dataset of the grid boundary polygons that are 100 meters by 100 meters. The file was converted to points via the grid polygon centroids. There were **82,748** grid features. See variable grid_keep~ for the total n which is **78,437.**

Filename: **grid_sites.csv**

Geography Type Prefix Table

Geography type	Geography Prefix
Pollen Sites	p
Grid Points	g

PROJECTED COORDINATE SYSTEM INFORMATION

This project study area is the entirety of New York City, NY. Thus, all of the data was converted to and processed in NAD 1983 StatePlane New York Long Island FIPS 3104 Feet.

NAD_1983_StatePlane_New_York_Long_Island_FIPS_3104_Feet

WKID: 2263 Authority: EPSG
Projection: Lambert_Conformal_Conic
False_Easting: 984250.0
False_Northing: 0.0
Central_Meridian: -74.0
Standard_Parallel_1: 40.66666666666666
Standard_Parallel_2: 41.03333333333333
Latitude_Of_Origin: 40.16666666666666
Linear Unit: Foot_US (0.3048006096012192)
Geographic Coordinate System: GCS_North_American_1983
Angular Unit: Degree (0.0174532925199433)
Prime Meridian: Greenwich (0.0)
Datum: D_North_American_1983
Spheroid: GRS_1980
Semimajor Axis: 6378137.0
Semiminor Axis: 6356752.314140356

Inverse Flattening: 298.257222101

VARIABLES

Following is a list of variables and their corresponding definitions.

Point Site (Pollen & Grid) IDs and coordinate information

Note: variables flagged with * are for pollen sites only and variables flagged with ~ are for grid points only.

siteid*

The pollen site unique identification number. Provided by Kate Weinberger

gid~

The grid point unique identification number. Created by BEH-GIS with the Python statement: "100000 + !LatFID100m! + 1"

longitude*

The pollen site longitude value in decimal degrees WGS84.

latitude*

The pollen site latitude value in decimal degrees WGS84.

xcoord

The pollen site x coordinate in NAD_1983_StatePlane_New_York_Long_Island_FIPS_3104_Feet WKID: 2263 Authority: EPSG (see Projected Coordinate System Information).

ycoord

The pollen site y coordinate in NAD_1983_StatePlane_New_York_Long_Island_FIPS_3104_Feet WKID: 2263 Authority: EPSG (see Projected Coordinate System Information).

grid_keep~

Many of the NYCCAS grid points fall outside of the NYC Boro Boundary feature and thus will not have usable Neighborhood Buffer Geographies. **Drop these if grid_keep == 0** (see Neighborhood Geographies Original Areas and land areas p. 7). The Grid analytical dataset is **78,437.** 4,311 records are dropped due to not being contained by boro boundary.

Point Site Elevation Variables

Method for assigning elevation data to geography point (g for grid point, p for pollen site, represented as -).

Esri tool: Extract Values to Points (http://resources.arcgis.com/en/help/main/10.1/index.html#//009z0000002t000000)

Interpolate values at the point locations (optional)

Specifies whether or not interpolation will be used.

Checked—The value of the cell will be calculated from the adjacent cells with valid values using bilinear interpolation. NoData values will be ignored in the interpolation unless all adjacent cells are NoData.

About the NYC Digital Elevation Model (DEM) 10 foot (cell resolution).

The NYC Digital Elevation Model (DEM) 10 foot (cell resolution) data layer was created by James Quinn of the Built Environment and Health Group (http://beh.columbia.edu/). James used the Topo to Raster (http://resources.arcgis.com/en/help/main/10.1/index.html#//009z0000006s000000) ArcToolbox tool. All of the input data was from NYCMap (contours,

elev pts, surface waters, shorelines). Which were available to the public back in the summer of 2006. James spent a lot of time cleaning up the shoreline layer. The data was received from Sarah Williams of Columbia University's Spatial Information and Design Lab (SIDL) (http://www.spatialinformationdesignlab.org/).

-_point_elev10ft

Digital elevation of geography point (g for grid point, p for pollen site, represented as -) in feet for NYC DEM 10 foot grid. If '-_point_elev10ft' is -9999 then that record should likely be dropped for grid analysis.

About the NYC Digital Elevation Model (DEM) 1 foot (cell resolution).

The NYC Digital Elevation Model (DEM) 1 foot (cell resolution) data layer was created for NYC from a 2010 Light Detection And Ranging (LIDAR) mission. NYC Open Data 1 foot LIDAR Derived Elevation Model (https://data.cityofnewyork.us/City-Government/1-foot-Digital-Elevation-Model-DEM-/dpc8-z3jc?category=City-Government&view_name=1-foot-Digital-Elevation-Model-DEM-)

-_point_elev01ft*

Digital elevation of geography point (g for grid point, p for pollen site, represented as -) in feet for NYC DEM 1 foot grid.

Point Site Nearness to Water Bodies Variables

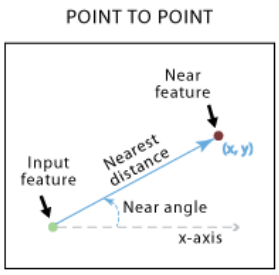
Method for assigning near water bodies to geography point (g for grid point, p for pollen site, represented as -)s

Esri Tool: Near Analysis (http://resources.arcgis.com/en/help/main/10.1/index.html#//000800000001q000000)

> Specifies whether the near angle values in decimal degrees will be calculated and written to a new field, NEAR_ANGLE. A near angle measures from the x-axis (horizontal axis) to the direction of the line connecting an input feature to its nearest feature at their closest locations, and it is within the range of 0 to 180 or 0 to -180 decimal degrees - 0 to the east, 90 to the north, 180 (-180°) to the west, and -90 to the south.

-_pointwaternearangle

The angle to the nearest water body. 0 to the east, 90 to the north, 180 (-180°) to the west, and -90 to the south



-_pointwaternearfeet

The distance to the nearest water body in feet. If 0 for water, then grid point is in or right on shoreline of water feature

-_pointwaternearmeters

The distance to the nearest water body in meters.

Neighborhood Geography Original Areas and Land Areas

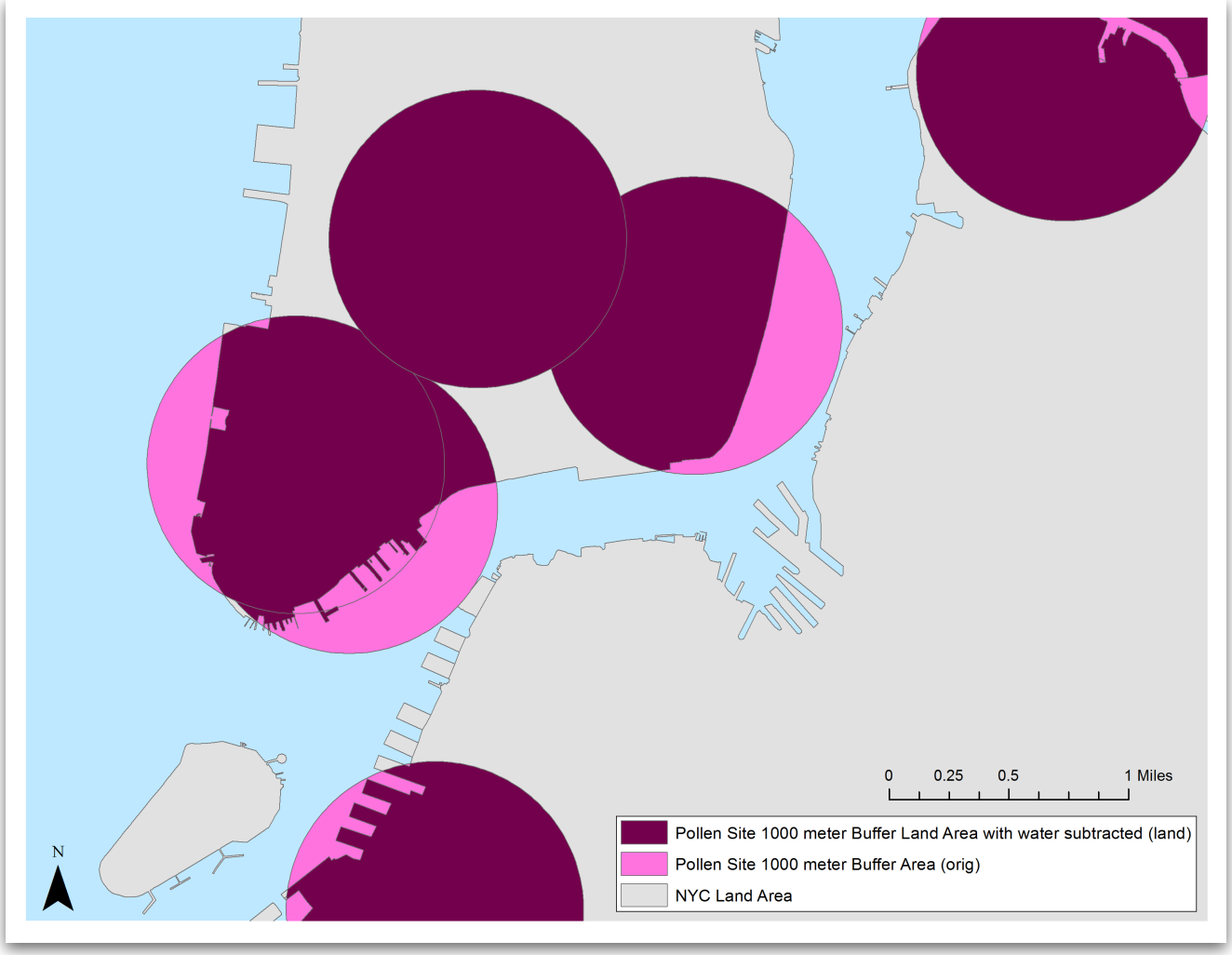
For this project the following neighborhood geography buffers were created (next page):

Each of these buffers was processed so that only land area of the point's residing borough was included. Buffers were allowed to cross boroughs if the borough boundaries were land based. Buffers with no portion containing a point site centroid were not created, and thus may be dropped for analytical purposes.

Neighborhood Geography Buffer Sizes List

Distance Kilometers	Distance Meters	Distance Feet	Buffer Type	Buffer Type Prefix	Distance Prefix
0.05 km	50 meters	164.042 feet	Radial	r	r0050m
0.1 km	100 meters	328.084 feet	Radial	r	r0100m
0.25 km	250 meters	820.210 feet	Radial	r	r0250m
0.5 km	500 meters	1640.420 feet	Radial	r	r0500m
1 km	1000 meters	3280.840 feet	Radial	r	r1000m

Original Buffer Area (orig) and Land Area (land) Diagram for Pollen Site 1000m (radial buffer)



Anatomy of a Variable Prefix

Geography prefix (anatomy of variable):

for **-** in Point Geographies List = [‘p’,’g'] (Pollen Sites, Grid Points)

for ^^^^^^ in Neighborhood Geography Buffers List = ['r0050m','r0100m','r0250m','r0500m','r1000m']

radial buffer - distance (meters)

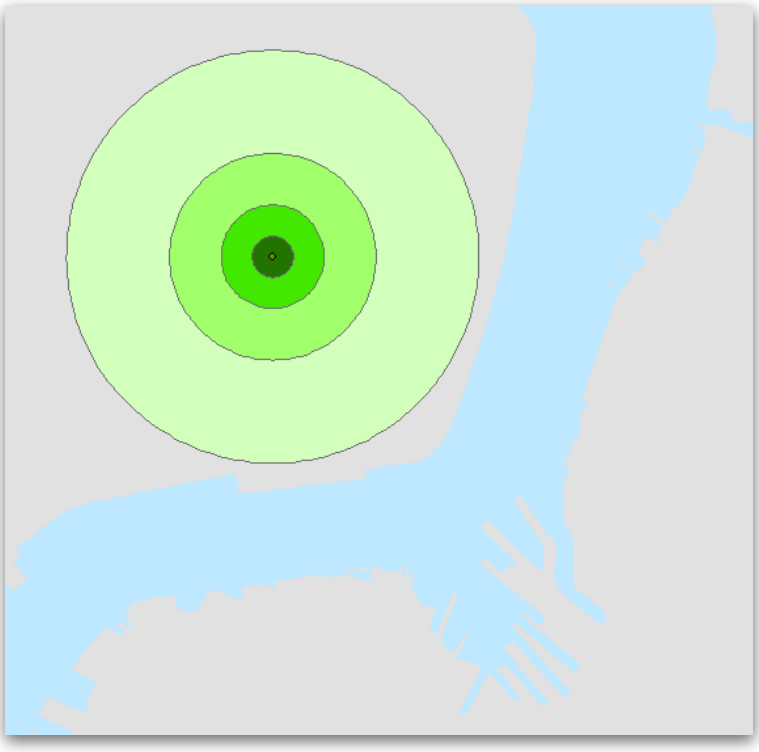
For example **pr0500mlandsqmtrs** is the variable for the pollen sites (p) radial buffer of 500 meters (r0500m) the neighborhood geography land area in meters (landsqmtrs). p + r0500m + landsqmtrs

-^^^^^^origsqmtrs

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography original area (including areas over water) in square meters.

-^^^^^^landsqmtrs

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography land area (water areas removed) in square meters.



50m, 100m, 250m, 500m 1000m buffers

Neighborhood Tree Canopy Area and Percentage Variables

Tree Canopy Diagram for Pollen Site 1000m (radial buffer)

Anatomy of a Variable Prefix

Geography prefix (anatomy of variable):

for **-** in Point Geographies List = [‘p’,’g'] (Pollen Sites, Grid Points)

for ^^^^^^ in Neighborhood Geography Buffers List = ['r0050m','r0100m','r0250m','r0500m','r1000m']

radial buffer - distance (meters)

For example **pr0500mtreecpctland** is the variable for the pollen sites (p) radial buffer of 500 meters (r0500m) percent of neighborhood geography land area that is tree canopy (treecpctland). p + r0500m + treecpctland

-^^^^^^treecsqmtrs

The amount of tree canopy area in the 50 meter buffer geography area in square meters.

-^^^^^^treecpctorig

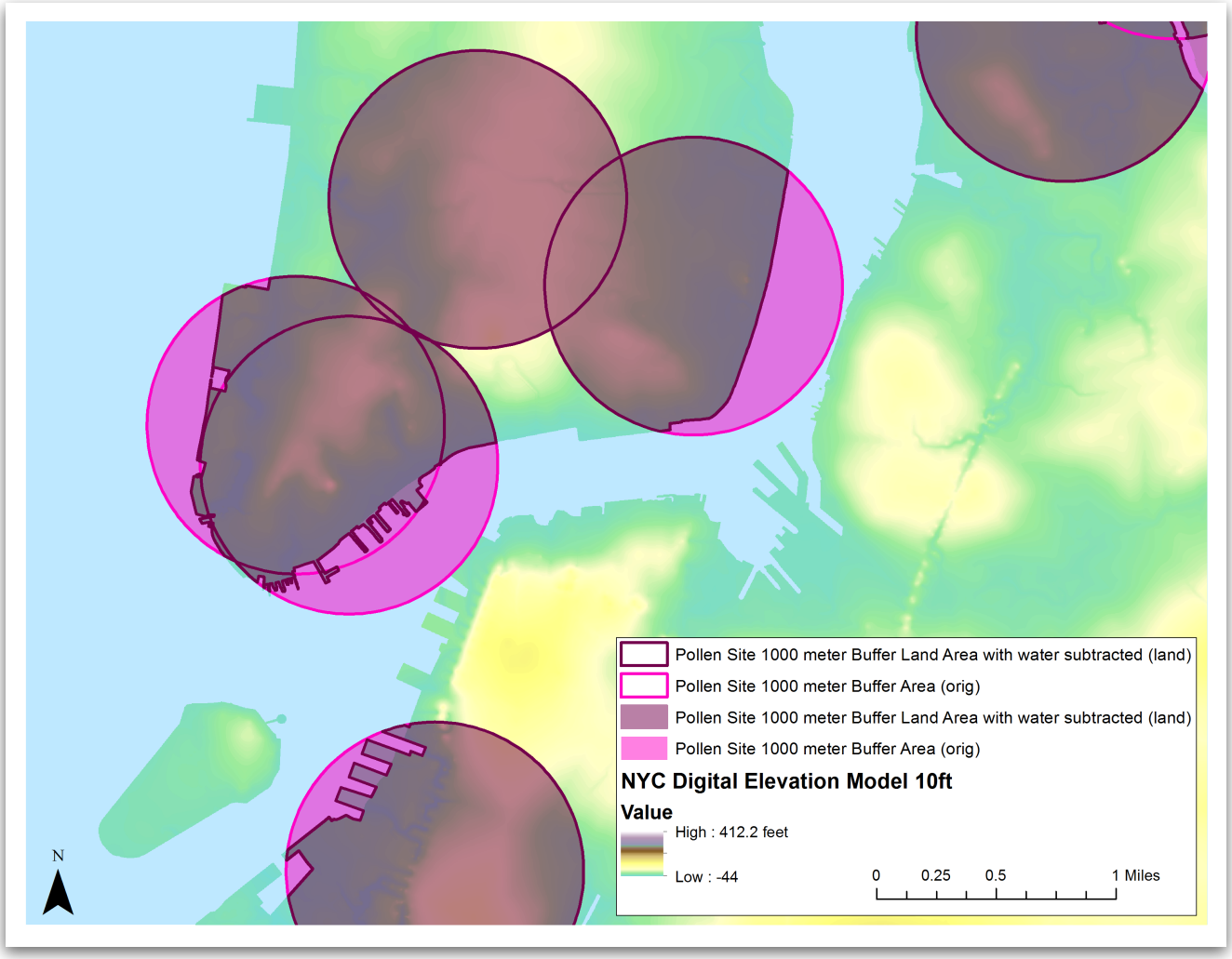
The percent of neighorhood area of the neighborhood geography 50 meter buffer original area (including areas over water) covered by tree canopy.

-^^^^^^treecpctland

The percent of neighorhood area of the neighborhood geography 50 meter buffer land area (water areas removed) covered by tree canopy.

Neighborhood Geography Elevation Variables

Elevation Diagram for Pollen Site 1000m (radial buffer)



Method: Zonal statistics as table (http://resources.arcgis.com/en/help/main/10.1/index.html#//009z000000w8000000) for land area.

Anatomy of a Variable Prefix

Geography prefix (anatomy of variable):

for **-** in Point Geographies List = ['p','g'] (Pollen Sites, Grid Points)

for ^^^^^^ in Neighborhood Geography Buffers List = ['r0050m','r0100m','r0250m','r0500m','r1000m']

radial buffer - distance (meters)

For example **pr0500melev10ftmean** is the variable for the pollen sites (p) radial buffer of 500 meters (r0500m) the 10 foot DEM elevation mean (elev10ftmean). p + r0500m + elev10ftmean

NYC DEM 10 FT Resolution Variables

The NYC Digital Elevation Model (DEM) 10 foot (cell resolution) data layer was created by [James Quinn](mailto:jq2145@columbia.edu?subject=10 Ft Elevation DEM") of the [Built Environment and Health Group](http://beh.columbia.edu/). James used the [Topo to Raster](http://resources.arcgis.com/en/help/main/10.1/index.html#//009z0000006s000000) ArcToolbox tool. All of the input data was from NYCMap (contours, elev pts, surface waters, shorelines). Which were available to the public back in the summer of 2006. James spent a lot of time cleaning up the shoreline layer. The data was received from Sarah Williams of [Columbia University's Spatial Information and Design Lab (SIDL)](http://www.spatialinformationdesignlab.org/)

-^^^^^^elev10ftcount

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) count of elevation raster grid cells contributing to the zonal statistics.

-^^^^^^elev10ftmin

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) minimum elevation raster grid cell contributing to the zonal statistics.

-^^^^^^elev10ftmax

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) maximum elevation raster grid cell contributing to the zonal statistics.

-^^^^^^elev10ftmean

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) mean elevation raster grid cell contributing to the zonal statistics.

-^^^^^^elev10ftstd

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) standard deviation of raster grid cells contributing to the zonal statistics.

NYC DEM 1 FT Resolution Variables

The NYC Digital Elevation Model (DEM) 1 foot (cell resolution) data layer was created for NYC from a 2010 Light Detection And Ranging (LIDAR) mission. [NYC Open Data 1 foot LIDAR Derived Elevation Model](https://data.cityofnewyork.us/City-Government/1-foot-Digital-Elevation-Model-DEM-/dpc8-z3jc?category=City-Government&view_name=1-foot-Digital-Elevation-Model-DEM-)

-^^^^^^elev01ftcount*

The geography point (p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) count of elevation raster grid cells contributing to the zonal statistics.

-^^^^^^elev01ftmin*

The geography point (p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) minimum elevation raster grid cell contributing to the zonal statistics.

-^^^^^^elev01ftmax*

The geography point (p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) maximum elevation raster grid cell contributing to the zonal statistics.

-^^^^^^elev01ftmean*

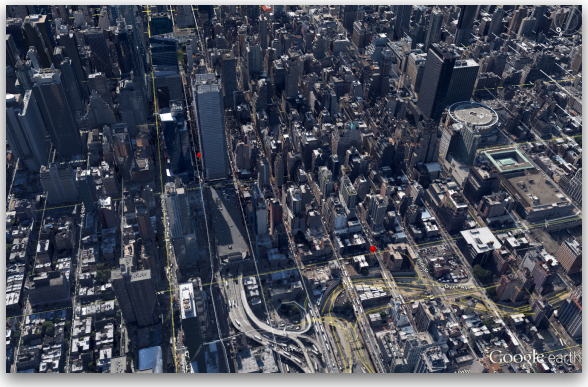
The geography point (p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) mean elevation raster grid cell contributing to the zonal statistics.

-^^^^^^elev01ftstd*

The geography point (p for pollen site, represented as -) neighborhood geography 50 meter buffer land area (water areas removed) standard deviation of raster grid cells contributing to the zonal statistics.

Building Variables

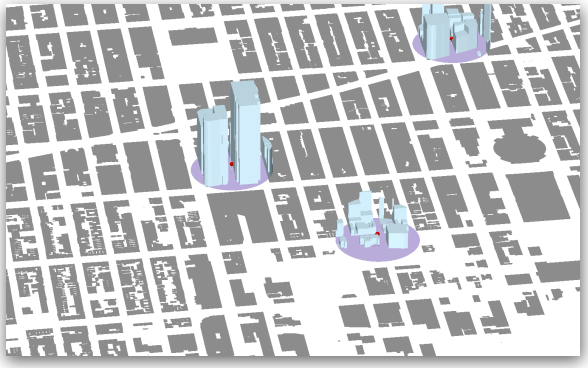
All building data was converted to meters. Volume is in square meters.



Building Volume

-^^^^^^bldgvol

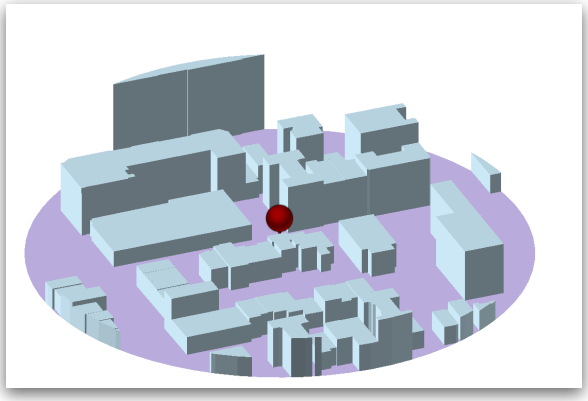
The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography (^^^^^^) buffer land area total building volume.



Building Bulk

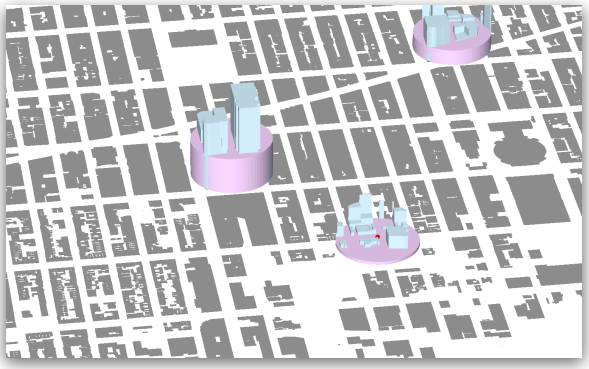
-^^^^^^bldgbulkorig

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography (^^^^^^) buffer land area total building volume/original area (including areas over water).



-^^^^^^bldgbulkland

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography (^^^^^^) buffer land area total building volume/land area (water areas removed).



Traffic Variables

The earliest birth date for children in the study cohort we're going to be estimating exposures for is in 2001, so I'd use that as the cutoff.

Anatomy of a Variable Prefix

Geography prefix (anatomy of variable):

for **-** in Point Geographies List = [‘p’,’g'] (Pollen Sites, Grid Points)

for ^^^^^^ in Neighborhood Geography Buffers List = ['r0050m','r0100m','r0250m','r0500m','r1000m']

radial buffer - distance (meters)

for **YYYY** in Years Range = (2001-2012) range of years for traffic obs. some missing due to no data intersected.

Esri Traffic Count Variables

The Esri Traffic data was for Years: 2001 - 2012. All 0 or null Traffic Count data was removed. The data was a point file (point - see traffic image below).

Annual Average Daily Traffic - 2001-2012

trafesri-^^^^^^avg_20012012

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography (^^^^^^) buffer land area (water areas removed) average Esri data traffic count for years: 2001 - 2012.

Annual Average Daily Traffic by Year - 2001-2012

trafesri-^^^^^^avg_YYYY

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography (^^^^^^) buffer land area (water areas removed) average Esri data traffic count by Year (YYYY). Years: 2001 - 2012.

New York State (NYS) DOT (Department of Transportation) Traffic Count Variables

NYS Traffic Data Viewer (http://gis.dot.ny.gov/tdv/) I emailed the NYS DOT and they sent the shapefile. The Year used for this project was 2012. The data is for street segments (line - see traffic image below).

trafnysd-^^^^^^avg_YYYY

The geography point (g for grid point, p for pollen site, represented as -) neighborhood geography (^^^^^^) buffer land area (water areas removed) average NYS DOT data traffic count by Year (YYYY).

Esri Traffic and NYS DOT Traffic Data Diagram for Pollen Site 1000m (radial buffer)