




# NIEHS Northern Manhattan SPARCs Time Estimate (v 1.0)

Prepared By:
Daniel M. Sheehan
Geographer, GIS Analyst
Built Environment and Health Project
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:
Dr. Gina Lovasi
Assistant Professor of Epidemiology
Built Environment and Health Project
Department of Epidemiology
Columbia University
722 West 168th Street, R804
New York, New York 10032
www.beh.columbia.edu
glovasi@columbia.edu

# TIME ESTIMATE INTRODUCTION

This time estimate represents the total number of hours required to construct each Type or Category of GIS metric(s) for ONE neighborhood definition (e.g., Census Tract, 1 radius buffer, etc.). However, constructing these metrics for TWO neighborhood definitions will not take double the time it takes to construct the metrics for ONE neighborhood definition because the initial set-up time to construct each metric is only required once, but not a second time. That is, there are initial tasks that need to be completed prior to constructing each metric (e.g., identifying, collecting, organizing, and pre-processing of input data, scripting of GIS methodologies, etc.), but once those initial task are complete for a metric, they do not need to be repeated a second time. As a result, please use the Neighborhood Definition Multiplier Table to determine the total number of hours required for constructing all of the metrics for n-number of neighborhood definitions.

A multiplier of 0.60 was applied to those metrics denoted with an "*" above their hourly estimate (i.e., Tasks: 1 – 18) in the Time Estimate for GIS Metrics table (page 3). When using the 0.60 multiplier, estimated hours in the Neighborhood Definition Multiplier Table were rounded up to the nearest whole number. However, work weeks (40-hours) and months were rounded to the nearest hundredth decimal place.

The context is a request for additional funds to the NIEHS Northern Manhattan center, building on the pilot projects that Kate and I, plus Diana Hernandez, have had with them. The original projects were $25K each, and this is a proposed add-on (perhaps in the 5-10K range). The ZIP-code plus 4 would be used to then link to SPARCs data for analyses.

## Target Time for Data

The target time to get data for this project is year 2011.

## Study Area

- Options are New York City (NYC) - only or NYC plus Nassau, Putnam, Rockland, Suffolk, and Westchester (NYC Metro).
- with respect to oil type, building bulk density, building age, tree canopy, Kate's to-be-created pollen surface (within NYC only) and
- land area, road classifications, sociodemographic characteristics and NETS business counts (all counties)
- targeting data that correspond to approximately 2011 (except NETS, which is 1990-2010)

# SPARCS

The following text is a description of SPARCS data from the SPARCS website (http://www.health.ny.gov/statistics/sparcs/access/):

> SPARCS is a comprehensive all payer data reporting system established in 1979 as a result of cooperation between the healthcare industry and government. The system was initially created to collect information on discharges from

hospitals. SPARCS currently collects patient level detail on patient characteristics, diagnoses and treatments, services, and charges for each hospital inpatient stay and outpatient (ambulatory surgery, emergency department, and outpatient services) visit; and each ambulatory surgery and outpatient services visit to a hospital extension clinic and diagnostic and treatment center licensed to provide ambulatory surgery services.

The enabling legislation for SPARCS is located under Section 28.16 of the Public Health Law (PHL). The regulations pertaining to SPARCS are under Section 400.18 of Title 10 (Health) of the Official Compilation of Codes, Rules, and Regulations of the State of New York (NYCRR).

## New York State ZIP code + 4 in NY State

The following text is a description of ZIP + 4 (aka ZIP plus 4) from Wikipedia (http://en.wikipedia.org/wiki/ZIP_code#ZIP.2B4):

In 1983, the U.S. Postal Service introduced an expanded ZIP code system that it called ZIP+4, often called "plus-four codes", "add-on codes", or "add ons". A ZIP+4 code uses the basic five-digit code plus four additional digits to identify a geographic segment within the five-digit delivery area, such as a city block, a group of apartments, an individual high-volume receiver of mail or any other unit that could use an extra identifier to aid in efficient mail sorting and delivery. But initial attempts to promote universal use of the new format met with public resistance,[citation needed] and today the plus-four code is not required. In general, mail is read by a multiline optical character reader (MLOCR) that almost instantly determines the correct ZIP+4 code from the address and—along with the even more specific delivery point—sprays an Intelligent Mail barcode (IM) on the face of the mail piece that corresponds to 11 digits—nine for the ZIP+4 code and two for the delivery point.

For Post Office Boxes, the general (but not invariable) rule is that each box has its own ZIP+4 code. The add-on code is often one of the following: the last four digits of the box number (e.g., PO Box 727050, Defreestville NY 12144-7050), zero plus the last three digits of the box number (e.g., PO Box 17727, Eagle River, AK 99577-0727), or, if the box number consists of fewer than four digits, enough zeros are attached to the front of the box number to produce a four-digit number (e.g., PO Box 77, Juneau AK 99750-0077). However, there is no uniform rule, so the ZIP +4 code must be looked up individually for each box.

It is common to use add-on code 9998 for mail addressed to the postmaster (to which requests for pictorial cancellations are usually addressed), 9999 for general delivery and other high-numbered add-on codes for business reply mail.[citation needed] For a unique ZIP code (explained below), the add-on code is typically 0001.

Not all addresses in the United States have had a Zip 4 code assigned to them. For those non-coded addresses, geocoding lookup or address validation which explicitly requires a Zip+4 may not succeed.

# Tasks Overview

## GEOCODING SOURCE DATA

In order to execute Geocoding the source data for the geocoding service must be developed.

### ZIP+4

#### Preliminary investigation into the ZIP+4 2000 TIGER/line data

There are 654,492 unique ZIP+4's from 2000 Census TIGER/line data for New York State. These points were mean center collapsed, meaning if multiple TIGER/line segments had the same ZIP+4, then these points were collapsed into a mean geographic center (unweighted) point. This processing was done in NAD 83 UTM Zone 18N (the preferred statewide projected coordinate system for NYS).

#### Esri 2010 Business Analyst Geocoder ZIP+4

Rather than building a ZIP+4 dataset for geocoding for 2010, the plan would be to use Esri's 2010 Business Analyst geocoder and geocode all of the unique SPARCs ZIP+4's and construct neighborhood geographies and generate variables from there. The United States Postal Service no longer sells ZIP+4 data with geographic coordinates (latitude,longitude) they only now publish a ZIP+4 list. There are other commercial vendors available but after installing ESRI 2010 Business Analyst it seems that this is

## GEOCODING

#### Circa 2000 SPARCs ZIP+4 data

Since the only geographic identifier for the SPARCs data is the ZIP+4 the geocoding will be relatively straight-forward. This process will be more like a table join, where the previously created 2000 ZIP+4 variables will be generated across NY State (or NYC depending on which options for the Study Area are selected) and then table-joined to the SPARCs list of circa 2000 ZIP+4's.

Unmatched ZIP+4's will be reviewed for possible typographical errors and/or corrected if possible, if not these records will be assigned 'missing' values.

#### Circa 2010 SPARCs ZIP+4 data

The 2010 data will follow a more conventional geocoding process using the Geocoder in Esri 2010 Business Analyst. ZIP+4's will be geocoded, unmatched records will be reviewed for possible typographical errors and/or corrected if possible, if not these records will be assigned 'missing' values.

# GEOPROCESSING TASKS FOR ALL NEIGHBORHOOD GEOGRAPHIES

## Geoprocessing tasks

Below are the tasks for which we only have usable and data ready in-hand for the New York City area.

1. Oil Burner Data (New York City (NYC)-only)
2. Pollen (New York City (NYC)-only)
3. NYCCAS Air Pollution PM2.5 and Black Carbon (New York City (NYC)-only)
4. Tree Canopy (New York City (NYC)-only)
5. Building Age (New York City (NYC)-only)
6. Building Bulk Density (New York City (NYC)-only)
7. NETS
8. 2009-2013 ACS 5-year estimates

# OPTIONS FOR STUDY AREA EXTENT AND VARIABLES:

There are four (4) options that Dr. Lovasi presented for generating a time estimate for:

1. New York Metro (NYC plus Nassau, Putnam, Rockland, Suffolk, and Westchester)
2. New York Metro without NETS data
3. New York City (only)
4. New York City (only) without NETS data

The Next section will explore each option and the corresponding time estimate.

## Tables for Study Area Extent and Summary by Task

**STUDY AREA EXTENT AND VARIABLES OPTIONS**

| | ZIP+4 circa 2000 count | Variable Multiplier |
|---|---|---|
| **New York State** | 654492 | - |
| **New York Metro (NYC with Nassau, Putnam, Rockland, Suffolk, and Westchester)*** | 302230 | 1.5 |
| **New York City*** | 103003 | 1.00 |

* Both include a 0.25 km buffer selection

## SUMMARY BY TASK WITH STUDY EXTENT AND VARIABLE OPTIONS FOR SINGLE GEOGRAPHY

| Category | New York City without NETS | New York City | New York Metro without NETS | New York Metro |
|---|---|---|---|---|
| **ZIP+4 Prep. Work** | 16h | 16h | 24h | 24h |
| **Geocoding** | 11h | 11h | 17h | 17h |
| **Neighborhood Geographies (orig,land (water-removed)** | 11h | 11h | *11h* | *11h* |
| **Oil Burner Data** | 8h | 8h | *8h* | *8h* |
| **Pollen** | 8h | 8h | *8h* | *8h* |
| **Air Pollution (BC & PM2.5)** | 16h | 16h | *16h* | *16h* |
| **Tree Canopy** | 16h | 16h | *16h* | *16h* |
| **Building Age** | 16h | 16h | *16h* | *16h* |
| **Building Bulk Density** | 16h | 16h | *16h* | *16h* |
| **NETS** | *0h* | 24h | *0h* | 36h |
| **2009-2013 ACS 5-year estimates** | 24h | 24h | 36h | 36h |
| **Total Hours** | **142h** | **166h** | **168h** | **204h** |
| **Total Days** | **18** | **21** | **21** | **25** |
| **Total Weeks** | **4** | **4** | **4** | **5** |

*UltraLight Italic cells mean that the task is either zero or unaffected by change in the Study Area extent*

# NEIGHBORHOOD GEOGRAPHIES

## ZIP+4 Geographies

ZIP+4 are points and therefore do not have an associated Polygon layer. It could be possible to create ZIP+4 catchment areas using Theissan Polygons (aka Voronoi diagrams). This would be considered as 1 geography.

## Buffer Geographies

Construct Radial Buffers for the following distances (in meters):

- 100m
- 250m
- 500m
- 1000m

Each of these will be considered a new geography and thus assigned the multiplier for each one used. Usual BEH protocol involves a **0.60** multiplier per Neighborhood Geography. Below is a Multiplier scale that assumes scalability of the geoprocessing so each Neighborhood Geography added doesn't require the same additional amount of time.

## Neighborhood Definition Multiplier Tables for Time

### NEIGHBORHOOD GEOGRAPHY MULTIPLIER BY HOURS

| | Multiplier | Non-multiplier hours | New York City without NETS | New York City | New York Metro without NETS | New York Metro |
|---|---|---|---|---|---|---|
| **1 Geography** | 1.0 | | 142h | 166h | 168h | 204h |
| **2 Geographies** | 1.5 | 27h | 146h | 182h | 164h | 218h |
| **3 Geographies** | 1.75 | 27h | 174h | 216h | 195h | 258h |
| **4 Geographies** | 1.95 | 41h | 184h | 231h | 207h | 277h |
| **5 Geographies** | 2 | 41h | 190h | 238h | 214h | 286h |

### NEIGHBORHOOD GEOGRAPHY MULTIPLIER BY WEEKS

| | New York City without NETS | New York City | New York Metro without NETS | New York Metro |
|---|---|---|---|---|
| **1 Geography** | 4 | 4 | 4 | 5 |
| **2 Geographies** | 4 | 5 | 4 | 5 |
| **3 Geographies** | 4 | 5 | 5 | 6 |
| **4 Geographies** | 5 | 6 | 5 | 7 |
| **5 Geographies** | 5 | 6 | 5 | 7 |

## Salary Bill Rate and Cost Summary Tables

From last time-estimate (2013) from Bijal Shah:

- The fringe is 31%, IC is 10%, # of hours per week is 35, and your hourly rate is $34.47.

### HOURLY RATE AND FRINGE AND INDIRECT COSTS

| Rate | Fringe | Indirect Costs | Hourly Rate with Fringe and IC |
|---|---|---|---|
| $ 34.67 | 0.31 | 0.1 | $ 48.88 |

### COST ESTIMATES BY COUNT OF NEIGHBORHOOD GEOGRAPHIES AND STUDY AREA VARIABLE OPTIONS

| | New York City w/out NETS | New York City | New York Metro w/out NETS | New York Metro |
|---|---|---|---|---|
| **1 Geography** | $ 6,941.63 | $ 8,114.86 | $ 8,188.19 | $ 9,948.04 |
| **2 Geographies** | $ 7,112.72 | $ 8,872.57 | $ 7,992.65 | $ 10,632.42 |
| **3 Geographies** | $ 8,518.16 | $ 10,571.32 | $ 9,544.74 | $ 12,624.47 |
| **4 Geographies** | $ 8,982.56 | $ 11,270.37 | $ 10,126.47 | $ 13,558.17 |
| **5 Geographies** | $ 9,263.65 | $ 11,610.12 | $ 10,436.88 | $ 13,956.58 |