National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Project

New York City Community Health Survey Dataset: 2002-2006

Variable Data Dictionary: Geospatial Measures of the Built and Social Environment

Prepared by:
James W. Quinn, MA
Geography, Senior GIS Analyst
Institute for Social and Economic Research and Policy
Built Environment and health Project
Columbia University
420 West 118th Street, MC 3355
New York, New York 10027
www.iserp.columbia.edu
www.beh.columbia.edu

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

AUGUST.23.2010

Table of Contents

NYC Community Health Survey (CHS)

The New York City Community Health Survey (CHS) is a telephone survey conducted annually by the DOHMH, Division of Epidemiology, Bureau of Epidemiology Services. CHS provides robust data on the health of New Yorkers, including neighborhood, borough and citywide estimates on a broad range of chronic diseases and behavioral risk factors.

CHS is based upon the National Behavioral Risk Factor Surveillance System (BRFSS), conducted by the Centers for Disease Control and Prevention. The CHS is a cross-sectional survey that samples approximately 10,000 adults aged 18 and older from all five boroughs of New York City – Manhattan, Brooklyn, Queens, Bronx, and Staten Island. A computer-assisted telephone interviewing (CATI) system is used to collect survey data, and interviews are conducted in a variety of different languages. All data collected are self-report.

The survey results are analyzed and disseminated in order to influence health program decisions, to increase the understanding of the relationship between health behavior and health status, and to support health policy positions.

The above text was copied directly from: http://www.nyc.gov/html/doh/html/survey/survey.shtml

CHS Weighted Variables: 2002-2006

♠ For more detailed information regarding the questions asked by the CHS in each respective year, please see the CHS Crosswalk Tables (i.e., DOE-CHS-2002-2006-Crosswalk-Table-Batch-1.xls and DOE-CHS-2002-2006-Crosswalk-Table-Batch-2.xls) that were provided as part of the final deliverables for this project.

The 10-character variable names in **BOLD** are the actual variable names in the GIS dataset (ArcGIS truncates variables names to 10-characters). The original CHS variable names are given in (parentheses) for the sake of reference and consistency.

+ Flags those variables that were delivered in *both* **Batch-1** and again in **Batch-2** of the 2002-06 CHS data variables.
^ Flags those variables that were delivered in *only* **Batch-2** of the 2002-06 CHS data variables.

+**CID** (Cid) – Unique CHS respondent identifier.

^**SURVEY** (Survey) – CHS survey year. Due to multiple waves in some years the survey year doesn't directly correspond.

POSSIBLE VALUES:
- 1 = 2002
- 3 = 2003
- 5 = 2004
- 6 = 2005
- 7 = 2006

^**DEMOG1** (Demog1) – **Continuous variable**. Age of CHS respondent.

^**DEMOG25** (Demog25) – Categorical age group. If a respondent did not know their age or refused to report their age for variable **DEMOG1**, they were asked, *"We are only asking this information to make sure that we have talked to enough people in each age group. Can you just tell me if you are:"*

POSSIBLE VALUES:
- 1 = '65 or older'
- 2 = '45-64'
- 3 = '24-44'
- 4 = '18-24'

+**SEX0206F** (Sex0206f) – Sex, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Male'
- 2 = 'Female'

+**NEWRACE020** (Newrace0206f) – Race/ethnicity, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'White'
- 2 = 'Black'
- 3 = 'Hispanic'
- 4 = 'Asian/Pacific Islander'
- 5 = 'Other' (includes multi-race, Native American, other)

+**MARITALSTA** (Maritalstatus0206f) – Marital Status, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Married'
- 2 = 'Divorced'
- 3 = 'Widowed'
- 4 = 'Separated'
- 5 = 'Never married'
- 6 = 'A member of unmarried couple'

+**EDUCATION0** (Education0206f) – Educational status, ages 25+, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Less than high school'
- 2 = 'High school graduate'
- 3 = 'Some college/technical school'
- 4 = 'College graduate'

+**EMPLOYMENT** (Employment0206f) – Employment status, response options differ in some years, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Employed for wages'
- 2 = 'Self-employed'
- 3 = 'Out of work >1 year'
- 4 = 'Out of work <1 year'
- 5 = 'Homemaker'
- 6 = 'Student'
- 7 = 'Retired'
- 8 = 'Unable to work'

+**NEWPOVGRPS** (Newpovgrps0206f) – Poverty group categories, income questions asked differently in 02-03 from 04-06, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = '<100%'
- 2 = '100-199%'
- 3 = '200-399%'
- 4 = '400-599%'
- 5 = '600%+'

+**CHILDREN02** (Children0206f) – Number children under 18 in household, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'none'
- 2 = 'One'
- 3 = 'Two'
- 4 = 'Three'
- 5 = 'More than three'

+**AGEGROUP02** (Agegroup0206f) – Categorical age group, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = '18 - 24'
- 2 = '25 - 44'
- 3 = '45 - 64'
- 4 = '65+'

+**AGE25UP020** (Age25up0206f) – Age 25 or older, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = '25 - 44'
- 2 = '45 - 64'
- 3 = '65+'

+**AGE18_6402** (Age18_640206f) – Age 18 - 64 only, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = '18 - 24'
- 2 = '25 - 44'
- 3 = '45 - 64'

+**AGE50UP020** (Age50up0206f) – Age 50 or older, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = '50 - 64'
- 2 = '65+'

^**PERSONHEIG** (Personheight) – Height of respondent in US inches.

^**PERSONWEIG** (Personweight) – Weight of respondent in US pounds.

+**WEIGHTALL0** (Weightall0206f) – Weight in 3 categories (persons with and without full data for BMI), all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Under/normal weight'
- 2 = 'Over weight'
- 3 = 'Obese'

+**BMI0206F** (Bmi0206f) – Body Mass Index (NOTE: set BMI upper limit =100). Wording changes, CHS 2002-06.

+**ZIPANALYZE** (Zipanalyze0206f) – Zip code variable to use for zip level analysis; includes some variables available <5 year (NOTE: some zips combined, CHS 2002-06). THIS VARIABLE WAS USED TO JOIN THE CHS VARIABLES TO GIS VARIABLES OF THE BUILT AND SOCIAL ENVIRONMENTS.

+**ZIPHOLD020** (Ziphold0206f) – NOT FOR ANALYSES - Zip code used for CHS weighting purposes, all CHS 2002-06.

+**ZIPWT0206F** (Zipwt0206f) – Zip code weight CHS 2002-06 all (created: 12.2008).

+**ZIPWT24P** (Zipwt24p) – Zip code weight CHS 2002, 2004 (created: 04.2009).

+**ZIPWT25P** (Zipwt25p) – Zip code weight CHS 2002, 2005 (created: 04.2009).

+**ZIPWT2346P** (Zipwt2346p) – Zip code weight CHS 2002, 2003, 2004, 2006 (created: 04.2009).

+**ZIPWT2345P** (Zipwt2345p) – Zip code weight CHS 2002, 2003, 2004, 2005 (created: 04.2009).

+**UHF420206F** (Uhf420206f) – 42 United Hospital Fund neighborhoods, all CHS 2002-06.

POSSIBLE VALUES:

- 1 = '101 Kingsbridge'
- 2 = '102 Northeast Bronx '
- 3 = '103 Fordham - Bronx Pk'
- 4 = '104 Pelham '
- 5 = '105 Crotona-Tremont'
- 6 = '106 High Bridge'
- 7 = '107 Hunts Point'
- 8 = '201 Greenpoint'
- 9 = '202 Downtown-Heights'
- 10 = '203 Bed-Stuy'
- 11 = '204 East New York'
- 12 = '205 Sunset Park'
- 13 = '206 Borough Park'
- 14 = '207 Flatbush'
- 15 = '208 Canarsie'
- 16 = '209 Bensonhurst'
- 17 = '210 Coney Island'
- 18 = '211 Williamsburg'
- 19 = '301 Washington Heights'
- 20 = '302 Central Harlem'
- 21 = '303 East Harlem'
- 22 = '304 Upper West Side'
- 23 = '305 Upper Eastside'
- 24 = '306 Chelsea-Clinton'
- 25 = '307 Gramercy Park'
- 26 = '308 Greenwich Village'
- 27 = '309 Union Square'
- 28 = '310 Lower Manhattan'
- 29 = '401 LIC, Astoria'
- 30 = '402 West Queens'
- 31 = '403 Flushing'
- 32 = '404 Bayside Little Neck'
- 33 = '405 Ridgewood'
- 34 = '406 Fresh Meadows'
- 35 = '407 South West Queens'
- 36 = '408 Jamaica'
- 37 = '409 South East Queens'
- 38 = '410 Rockaway'
- 39 = '501 Port Richmond'
- 40 = '502 Stapleton'
- 41 = '503 Willbrook'
- 42 = '504 South Beach'

+**BOROUGH020** (Borough0206f) – Borough of residence, all CHS 2002-06.

POSSIBLE VALUES:

- 1 = 'Bronx'
- 2 = 'Brooklyn'
- 3 = 'Manhattan'
- 4 = 'Queens'
- 5 = 'Staten Island'

+**DPHONEW020** (Dphonew0206f) – District Public Health Offices, (NOTE: boundaries use 2006 definitions, all CHS 2002-06).

POSSIBLE VALUES:

- 1 = 'South Bronx'
- 2 = 'East and Central Harlem'
- 3 = 'North and Central Brooklyn (change in 2006)'
- 4 = 'All Other Neighborhoods'

^**SAFENEIGH1** (Safeneigh1p) – How safe from crime do you consider your neighborhood to be?

POSSIBLE VALUES:

- 1 = 'Extremely safe'
- 2 = 'Quite safe'
- 3 = 'Slightly safe'
- 4 = 'Not at all safe'

+**FRUITVEG24** (Fruitveg24p) – How many total servings of fruit and/or vegetables did you eat yesterday? Categorical Variable: CHS 2002, 2004.

POSSIBLE VALUES:
- 1 = 'none'
- 2 = '1-4'
- 3 = '5+'

+**FRUITVEG_C** (Fruitveg_C24p) – **Continuous variable**. How many total servings of fruit and/or vegetables did you eat yesterday? Continuous Variable: CHS 2002, 2004.

^**BUYFVNEIGH** (Buyfvneigh1p) – When you or someone in your household shops for fresh fruits or vegetables, do you buy them in your neighborhood?

POSSIBLE VALUES:
- 1 = 'Yes, in my neighborhood'
- 2 = 'No, someplace else'

^**REASONFV1P** (Reasonfv1p) – What is the main reason you or they don't buy fresh fruits and vegetables in your neighborhood?

POSSIBLE VALUES:
- 1 = 'No stores'
- 2 = 'Poor quality'
- 3 = 'Expensive'
- 4 = 'Poor service'
- 5 = 'Uncomfortable'
- 6 = 'Do not cook'
- 7 = 'Do not eat fruits and vegetables'
- 8 = 'Buys in bulk'
- 9 = 'Buys near job'
- 10 = 'Buys organic food only'
- 11 = 'Do not like neighborhood stores'
- 12 = 'Other'

^**ACTIVEWORK** (Activework3p) – When you are at work, which of the following best describes what you do?

POSSIBLE VALUES:
- 1 = 'Mostly sitting or standing'
- 2 = 'Mostly walking'
- 3 = 'Mostly heavy labor/physically demanding work'
- 4 = 'Not currently working'

^**SODAPERDAY** (Sodaperday6p) – On an average day, how many sodas do you drink? One drink of soda would equal a 12 ounce can, bottle, or glass. Do not include diet soda or seltzer.

POSSIBLE VALUES:
- 1 = 'none'
- 2 = '1 12oz'
- 3 = '2+ 12 oz'

^**SODAPERDA0** (Sodaperday_C6p) – **Continuous variable**. On an average day, how **many** sodas do you drink? One drink of soda would equal a 12 ounce can, bottle, or glass. Do not include diet soda or seltzer.

^**MEALSOUT6P** (Mealsout6p) – During an average week, how **many** times do you eat meals that were purchased at a restaurant, deli, or street vendor? Please include meals eaten at a restaurant, carried out or delivered.

POSSIBLE VALUES:
- 1 = 'None'
- 2 = '1-2 per week'
- 3 = '3-7 per week'
- 4 = '8+ per week'

^**MEALSOUT_C** (Mealsout_C6p) – **Continuous variable**. During an average week, how many times do you eat meals that were purchased at a restaurant, deli, or street vendor? Please include meals eaten at a restaurant, carried out or delivered.

^**COMPUTER6P** (Computer6p) – On an average day, how many **hours** do you spend using a computer for personal email, searching the internet, or playing games? DO not include time spent using a computer at work or for schoolwork.

POSSIBLE VALUES:
- 1 = 'None'
- 2 = '< half hr'
- 3 = '1 hr'
- 4 = '2 hrs'
- 5 = '>= 3 hrs'

^**COMPUTER_C** (Computer_C6p) – **Continuous variable: Hours**. On an average day, how many **hours** do you spend using a computer for personal email, searching the internet, or playing games? DO not include time spent using a computer at work or for schoolwork.

** Refer to questionnaire for continuous values for computer use. Values of 88 refer to <30 minutes.*

^**USEWALKTRA** (Usewalktrail6p) – Do you use walking trails, parks, playgrounds, or sports fields in your neighborhood for physical activity? Would you say yes, no or my neighborhood does not have these facilities?

POSSIBLE VALUES:
- 1 = 'Yes'
- 2 = 'No'
- 3 = 'My neighborhood does not have these facilities'

^**SAFEWALKTR** (Safewalktrail6p) – How safe are the walking trails, parks, playgrounds, and sports fields in your neighborhood? Would you say it is very safe, somewhat safe, somewhat unsafe, or very unsafe?

POSSIBLE VALUES:
- 1 = 'Very safe'
- 2 = 'Somewhat safe'
- 3 = 'Somewhat unsafe'
- 4 = 'Very unsafe'

^**TENBLOCKS3** (Tenblocks35p) – Over the past 30 days, have you walked or bicycled more than 10 blocks as part of getting to and from work, or school, or to do errands?

POSSIBLE VALUES:
- 1 = 'Yes'
- 2 = 'No'
- 3 = 'Unable to do activity'

^**TENBLOCKSD** (Tenblocksdays_C3p) – **Continuous variable: Days**. Number of **days** walked/biked 10-blocks. Over the past 30 days, how often did you do this? How many times per day, per week, or per month did you do these activities?

^**TENBLOCKSW** (Tenblocksweeks_C3p) – **Continuous variable: Weeks**. Number of **weeks** walked/biked 10-blocks. Over the past 30 days, how often did you do this? How many times per day, per week, or per month did you do these activities?

^**TENBLOCKSM** (Tenblocksmonths_C3p) – **Continuous variable: Months**. Number of **months** walked/biked 10-blocks. Over the past 30 days, how often did you do this? How many times per day, per week, or per month did you do these activities?

^**EXPERWEEK2** (Experweek20min1p) – On average, how many times during the week do you exercise for at least 20 minutes each time?

POSSIBLE VALUES:
- 1 = '< 1 time per week'
- 2 = '1 time'
- 3 = '2 times'
- 4 = '3 times'
- 5 = '>3 times'
- 6 = 'None'

^**EXPERWEEK3** (Experweek30min3p) – On average, how many **days** per week do you exercise at least 30 minutes?

POSSIBLE VALUES:
- 1 = 'None'
- 2 = '1 - 2 days'
- 3 = '3 - 4 days'
- 4 = '5 - 7 days'

^**EXPERWEEK0** (Experweek30min_C3p) – **Continuous variable: Days**. On average, how many **days** per week do you exercise at least 30 minutes?

+**EXERCISE23** (Exercise2345p) – During past 30 days, other than regular job, did you participate in any physical activities or exercises? Wording changes, CHS 2002-05.

POSSIBLE VALUES:
- 1 = 'Yes'
- 2 = 'No'

^**VIGFREQDAY** (Vigfreqdays67p) – **Continuous variable: Day**. How many times per **day** do you do vigorous physical activities for at least 10 minutes that cause heavy sweating or large increase in breathing or heart rate?

^**VIGFREQWEE** (Vigfreqweek67p) – **Continuous variable: Week**. How many times per **week** do you do vigorous physical activities for at least 10 minutes that cause heavy sweating or large increase in breathing or heart rate?

^**VIGFREQMNT** (Vigfreqmnth67p) – **Continuous variable: Month**. How many times per **month** do you do vigorous physical activities for at least 10 minutes that cause heavy sweating or large increase in breathing or heart rate?

^**VIGTIMEMIN** (Vigtimemin67p) – **Continuous variable: Minutes**. About how many **minutes** do you do these vigorous physical activities each time?

^**VIGTIMEHRS** (Vigtimehrs67p) – **Continuous variable: Hours**. About how many **hours** do you do these vigorous physical activities each time?

^**MODFREQDAY** (Modfreqdays67p) – **Continuous variable: Day**. How often do you do light or moderate physical activities for at least 10 minutes that cause only light sweating or a slight to moderate increase in breathing or heart rate?

^**MODFREQWEE** (Modfreqweek67p) – **Continuous variable: Week**. How often do you do light or moderate physical activities for at least 10 minutes that cause only light sweating or a slight to moderate increase in breathing or heart rate?

^**MODFREQMNT** (Modfreqmnth67p) – **Continuous variable: Month**. How often do you do light or moderate physical activities for at least 10 minutes that cause only light sweating or a slight to moderate increase in breathing or heart rate?

^**MODTIMEMIN** (Modtimemin67p) – **Continuous variable: Minutes**. About how long do you do these light to moderate physical activities each time?

^**MODTIMEHRS** (Modtimehrs67p) – **Continuous variable: Hours**. About how long do you do these light to moderate physical activities each time?

^**WORKSCHZIP** (Workschzip7p) – What is the zip code of your work or school where you spend the most time from 9 to 5 on weekdays?

^**WORKSCHINT** (Workschinta7p) – What is the nearest intersection or the cross streets of your place of work or school where you spend most of your time from 9 to 5 weekdays?

^**WORKSCHIN0** (Workschintb7p) – What is the nearest intersection or the cross streets of your place of work or school where you spend most of your time from 9 to 5 weekdays?

^**WORKSCHBOR** (Workschboro7p) – What is the nearest intersection or the cross streets of your place of work or school where you spend most of your time from 9 to 5 weekdays?

POSSIBLE VALUES:
- 1 = 'Bronx'
- 2 = 'Brooklyn'
- 3 = 'Manhattan'
- 4 = 'Queens'
- 5 = 'Staten Island'
- 6 = 'Other, not in NYC'

+**DIABETES23** (Diabetes2346p) – Ever been told by a doctor that you have diabetes? Wording changes, CHS 2002, 2003, 2004, 2006.

POSSIBLE VALUES:
- 1 = 'Yes'
- 2 = 'No'

+**GENERALHEA** (Generalhealth0206f) – Self-reported general health status, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Excellent'
- 2 = 'Very Good'
- 3 = 'Good'
- 4 = 'Fair'
- 5 = 'Poor'

+**PCP0206F** (Pcp0206f) – 1+ primary care provider, wording and response options differ between years, all CHS 2002-06

POSSIBLE VALUES:
- 1 = 'Yes'
- 2 = 'No'

+**INSURED020** (Insured0206f) – Insured/uninsured - created from type insurance-questions asked differently 02-03 from 04-06, all CHS 2002-06

POSSIBLE VALUES:
- 1 = 'Yes'
- 2 = 'No'

+**FLUSHOT020** (Flushot0206f) – Flu shot/spray in past 12 months, ages 50+, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Yes'
- 2 = 'No'

+**SMOKER0206** (Smoker0206f) – Smoking status, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Never'
- 2 = 'Current'
- 3 = 'Former'

+**EVERYDAY02** (Everyday0206f) – Smoke every day vs. some days (yes/no), all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Yes (Every day)'
- 2 = 'No (Some days)'
- 3 = 'Not at all'

+**CIGBUYSIX0** (Cigbuysix0206f) – Where did you get last cigarette (6 category version), response options differ in some years, all CHS 2002-06.

POSSIBLE VALUES:
- 1='New York City'
- 2='Outside NYC but in NYS'
- 3='Internet/mail'
- 4='Another person/street location unknown'
- 5='Indian reservation, different state, duty-free, outside US'
- 6='Other'

+**CONDOM1864** (Condom18640206f) – Condom use at last sex, 18-64 years only, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'Yes'
- 2 = 'No'

+**SEXPARTNER** (Sexpartner18640206f) – Number sex partners in the past 12 months, 18-64 years only, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'None'
- 2 = 'One'
- 3 = 'Two'
- 4 = 'Three or more'

^**YEARSINUS1** (Yearsinus136p) – Since you moved to the United States, how long have you lived in this country?

Two variables (not included with this dataset) with difference response categories were standardized and combined to create this variable: ***DEMOG14*** *(2002+2005) and* ***DEMOG26*** *(2003). In regards to analysis this variable can be treated as the same in 2002, 2003, and 2005.*

POSSIBLE VALUES:
- 1 = '< 1 year'
- 2 = '1 to <4 years'
- 3 = '4+ years'

^**HOWLONGINU** (Howlonginus7p) – Since you moved to the United States, how long have you lived in this country?

The response categories for this variable are different than those in 2002, 2003, and 2005. In regards to analysis this variable ***cannot*** *be used with the* ***YEARSINUS1*** *variable because the response categories are different*

POSSIBLE VALUES:
- 1 = '< 5 years'
- 2 = '5 - 9 years'
- 3 = '10+ years'

+**USBORN0206** (Usborn0206f) – Nativity - US or Foreign born, all CHS 2002-06.

POSSIBLE VALUES:
- 1 = 'US born'
- 2 = 'Foreign born'

+**BTHREGION0** (Bthregion0206f) – Place of birth, includes regions, all CHS 2002-06.

POSSIBLE VALUES:
- Please see Appendix A for individual County codes.

+**COUNTRYBIR** (Countrybirth0206f) – Country of Birth, (regions NOT included), all CHS 2002-06.

POSSIBLE VALUES:
- Please see Appendix A for individual County codes.

^**DADBTHREGI** (Dadbthregion6p) – In what country was your father born (includes regions).

POSSIBLE VALUES:
- Please see Appendix A for individual County codes.

^**DADCOUNTRY** (Dadcountrybirth6p) – In what country was your father born?

POSSIBLE VALUES:
- Please see Appendix A for individual County codes.

^**MOMBTHREGI** (Mombthregion6p) – In what country was your mother born (includes regions).

POSSIBLE VALUES:
- Please see Appendix A for individual County codes.

^**MOMCOUNTRY** (Momcountrybirth6p) – In what country your mother born?

POSSIBLE VALUES:
- Please see Appendix A for individual County codes.

^**ATHOMELANG** (Athomelang567p) – What language do you speak most often at home?

POSSIBLE VALUES:
- 1 = 'English'
- 2 = 'Spanish'
- 3 = 'Other'

^**ATHOMELAN0** (Athomelang67p) – What language do you speak most often at home?

POSSIBLE VALUES:
- 1 = 'English'
- 2 = 'Spanish'
- 3 = 'Russian'
- 4 = 'Chinese'
- 5 = 'Indian'
- 6 = 'Other'

GIS Measures of the Built and Social Environment

Below is a list and brief description of the social and built environment variables that were calculated for the CHS Zip Codes. Please use this document as a data dictionary for the accompanying tables.

ZIP – Zip Code variable used for GIS processing and unique table joining feature key ID.
ZIP_KM2 – Total area of Zip Code in Km2.
ZIP_LNDKM2 – Total land area of Zip Code in Km2 (inland water features subtracted out). THIS VARIABLE SHOULD BE USED WHEN CALCULATING DENSITY ESTIMATION VARIABLES.

Social Environment Variables

US Census Summary File 3 Variables

♠ ALL 2000 US CENSUS VARIABLES WERE CALCULATED USING VARIABLES FROM SUMMARY FILE 3 [SF3] SAMPLE DATA <www.census.gov/prod/cen2000/doc/sf3.pdf>. WHEN POSSIBLE, VARIABLES WERE FIRST CALCULATED USING BLOCK GROUP LEVEL VARIABLES. HOWEVER, THERE ARE CENSUS VARIABLES AS PART OF THIS GEO-SPATIAL ANALYSIS, WHICH WERE ONLY AVAILABLE AT THE CENSUS TRACT LEVEL [I.E., LANGUAGE SPOKEN AT HOME VARIABLES].

Age Variables

ZIP_TOTPOP – Total population [P001001].
ZIP_POPDEN – Population density per Km2 [P001001] / [ZIP_LNDKM2].
ZIP_TOTMAL – Total population: Male [P008002].
ZIP_TOTFEM – Total population: Female [P008041].

Population & Race Variables

ZIP_CNTWHT – Count of the White population (from the 'single race' variable) [P006002].
ZIP_PCTWHT – Percent White population (from the 'single race' variable) [P006002] / [P006001].
ZIP_CNTBLK – Count of the Black or African American population (from the 'single race' variable) [P006003].
ZIP_PCTBLK – Percent Black or African American population [P006003] / [P006001].
ZIP_CNTLAT – Count of the Hispanic or Latino population (from Hispanic or Latino by Race variable) [P007010]
ZIP_PCTLAT – Percent of the Hispanic or Latino population [P007010] / [P007001].
ZIP_CNTNAT – Count of the American Indian or Alaska Native (includes: American Indian, American Indian tribe, Alaska Native) population (from the 'single race variable) [P006004].

ZIP_PCTNAT – Percent American Indian or Alaska Native (includes: American Indian, American Indian tribe, Alaska Native) population [P006004] / [P006001].
ZIP_CNTASN – Count of the Asian (includes: Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, Other Asian) population (from the 'single race variable) [P006005].
ZIP_PCTASN – Percent of the Asian (includes: Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, Other Asian) population [P006005] / [P006001].
ZIP_CNTPAC – Count of the Native Hawaiian or Other Pacific Islander (includes: Native Hawaiian, Guamanian, Chamorro, Samoan, Other Pacific Islander—Carolinian ; Chuukese (Trukese); Fijian; Kosraean; Melanesian; Micronesian; Northern Mariana Islander; Palauan; Papua New Guinean; Pohnpeian; Polynesian; Solomon Islander; Tahitian; Tokelauan; Tongan; Yapese) population (from the 'single race' variable) [P006006].
ZIP_PCTPAC – Percent of the Native Hawaiian or Other Pacific Islander (includes: Native Hawaiian, Guamanian, Chamorro, Samoan, Other Pacific Islander—Carolinian ; Chuukese (Trukese); Fijian; Kosraean; Melanesian; Micronesian; Northern Mariana Islander; Palauan; Papua New Guinean; Pohnpeian; Polynesian; Solomon Islander; Tahitian; Tokelauan; Tongan; Yapese) population [P006006] / [P006001].
ZIP_CNTOTR – Count of the Some other race population (from the 'single race' variable) [P006007].
ZIP_PCTOTR – Percent of the Some other race population [P006007] / [P006001].
ZIP_CNTTWO – Count of the Two or more races population (from the 'single race' variable) [P006008].
ZIP_PCTTWO – Percent of the Two or more races population [P006008] / [P006001].
ZIP_FORBRN – Total foreign born population [P021013].
ZIP_FORDEN – Density of foreign born population per Km2 [P021013 / ZIP_LNDKM2].
ZIP_PCTFOR – Percent of foreign born population [P021013] / [P001001]].
ZIP_PCTPOV – Percent of total population for whom poverty status is determined [no institutionalized population] whose income is below 100% of the federal poverty line [((P088002 + P088003 + P088004) / P088001)].
ZIP_MHHINC – Median household income in 1999 [dollars] from US Census sf3 [P053001]. Unlike most other US Census variables, this variable is not a count value but rather a median. Therefore, rather than apportioning the Census Block Groups where they are cut the ZIP Code measurement geographies, the median household income value for every Block Group that intersected each ZIP was considered and the median value of all those considered values was calculated.

Household Variables

ZIP_HHTOT – House-holds: Total [P020001].
ZIP_HHSOC – House-holds: With Social Security income [P062002].
ZIP_HHSSI – House-holds: With Supplemental Security Income (SSI) [P063002].
ZIP_HHPAI – House-holds: With public assistance income [P064002].

Housing Unit Variables

ZIP_HUTOT – Housing units: Total [H006001].
ZIP_HUOCCP – Occupied housing units: Total [H006002].
ZIP_PCTOCP – Percent of housing units that are occupied [H006002 / H006001].
ZIP_VACANT – Vacant housing units: Total [H006003].
ZIP_PCTVCT – Percent of house units that are vacant [H006003 / H006001].
ZIP_HUOWN – Owner occupied housing units: Total [H007002].
ZIP_PCTOWN – Percent of occupied house-holds that are owned [H007002 / H007001].
ZIP_HURENT – Renter occupied housing units: Total [H007003].
ZIP_PCTRNT – Percent of occupied house units that are rented [H007003 / H007001].

Commute to Work Variables

ZIP_PCTCAR – Percent of workers 16 years & over whom commuted to work using a private vehicle. Means of a private vehicle include cars; trucks; vans; carpooling; and driving alone [P030002 / P030001].
ZIP_PCTPUB – Percent of workers 16 years & older who commuted to work using public transportation. Means of public transportation include, bus or trolley bus; streetcar or trolley car; subway or elevated; railroad; ferryboat; and taxicab [P030005 / P030001].
ZIP_PCTWLK – Percent of workers 16 years & older who commuted to work by walking [P0300014/ P030001].
ZIP_PCTBIK – Percent of workers 16 years & older who commuted to work using a bicycle [P030013 / P030001].

Work Related Variables

The 'Work Related Variables' (P26, P27, and P30-P35) were added as a result of a discussion during the BEH Team Meeting on Tuesday, August 10th about how far people travel to get to work and by what means of transportation. These variables were identified as being relevant to that discussion and are all simple count variables representing the total number of workers (i.e., persons) or travel time to work in **minutes**. No percentage (%) variables were calculated for the simple reason that the numbers of variables presented in this section were many and it was unknown which variables would be of interest for analytical purposes. However, percentage variables can easily be calculated using the variable of interest as the numerator and the **total** population value from the respective census category as the denominator. E.G., to calculate the 'percent of workers who worked in their county of residence' you would use the following expression: (P026003 / P026001]. Variables under the P30 category in red are flagged as they were previously used to calculate several variables (ZIP_PCTCAR, ZIP_PCTPUB, ZIP_PCTWLK, and ZIP_PCTBIK).

P26. Place of Work for Workers 16 Years and Over--State and County Level

Universe: Workers 16 years and over

Total:	**P026001**
Worked in state of residence:	P026002
Worked in county of residence	P026003
Worked outside county of residence	P026004
Worked outside state of residence	P026005

P27. Place of Work for Workers 16 Years and Over--Place Level

Universe: Workers 16 years and over

Total:	**P027001**
Living in a place:	P027002
Worked in place of residence	P027003
Worked outside place of residence	P027004
Not living in a place	P027005

P30. Means of Transportation to Work for Workers 16 Years and Over

Universe: Workers 16 years and over

Total:	**P030001**
Car, truck, or van:	P030002
Drove alone	P030003
Carpooled	P030004
Public transportation:	P030005
Bus or trolley bus	P030006
Streetcar or trolley car	P030007
Subway or elevated	P030008
Railroad	P030009
Ferryboat	P030010
Taxicab	P030011
Motorcycle	P030012
Bicycle	P030013
Walked	P030014
Other means	P030015
Worked at home	P030016

P31. Travel Time to Work for Workers 16 Years and Over

Universe: Workers 16 years and over

Total:	**P031001**
Did not work at home:	P031002
Less than 5 minutes	P031003
5 to 9 minutes	P031004
10 to 14 minutes	P031005
15 to 19 minutes	P031006
20 to 24 minutes	P031007
25 to 29 minutes	P031008
30 to 34 minutes	P031009
35 to 39 minutes	P031010
40 to 44 minutes	P031011
45 to 59 minutes	P031012
60 to 89 minutes	P031013
90 or more minutes	P031014
Worked at home	P031015

P32. Travel Time to Work by Means of Transportation to Work for Workers 16 Years and Over Who Did Not Work at Home

Universe: Workers 16 years and over who did not work at home

Total:	**P032001**
Less than 30 minutes:	P032002
Public transportation	P032003
Other means	P032004
30 to 44 minutes:	P032005
Public transportation	P032006
Other means	P032007
45 to 59 minutes:	P032008
Public transportation	P032009
Other means	P032010
60 or more minutes:	P032011
Public transportation	P032012
Other means	P032013

P33. Aggregate Travel Time to Work (in minutes) by Travel Time to Work by Means of Transportation to Work for Workers 16 Years and Over Who Did Not Work at Home

Universe: Workers 16 years and over who did not work at home

Aggregate travel time to work (in minutes)**:**	**P033001**
Less than 30 minutes:	P033002
Public transportation	P033003
Other means	P033004
30 to 44 minutes:	P033005
Public transportation	P033006
Other means	P033007
45 to 59 minutes:	P033008
Public transportation	P033009
Other means	P033010
60 or more minutes:	P033011
Public transportation	P033012
Other means	P033013

P34. Time Leaving Home to Go to Work for Workers 16 Years and Over

Universe: Workers 16 years and over

Total:	**P034001**
Did not work at home:	P034002
12:00 a.m. to 4:59 a.m.	P034003
5:00 a.m. to 5:29 a.m.	P034004
5:30 a.m. to 5:59 a.m.	P034005
6:00 a.m. to 6:29 a.m.	P034006
6:30 a.m. to 6:59 a.m.	P034007
7:00 a.m. to 7:29 a.m.	P034008
7:30 a.m. to 7:59 a.m.	P034009
8:00 a.m. to 8:29 a.m.	P034010
8:30 a.m. to 8:59 a.m.	P034011
9:00 a.m. to 9:59 a.m.	P034012
10:00 a.m. to 10:59 a.m.	P034013
11:00 a.m. to 11:59 a.m.	P034014
12:00 p.m. to 3:59 p.m.	P034015
4:00 p.m. to 11:59 p.m.	P034016
Worked at home	P034017

P35. Private Vehicle Occupancy for Workers 16 Years and Over

Universe: Workers 16 years and over

Total:	**P035001**
Car, truck, or van:	P035002
Drove alone	P035003
Carpooled:	P035004
In 2-person carpool	P035005
In 3-person carpool	P035006
In 4-person carpool	P035007
In 5- or 6-person carpool	P035008
In 7-or-more-person carpool	P035009
Other means (including those who worked at home)	P035010

Linguistic Isolation Variables

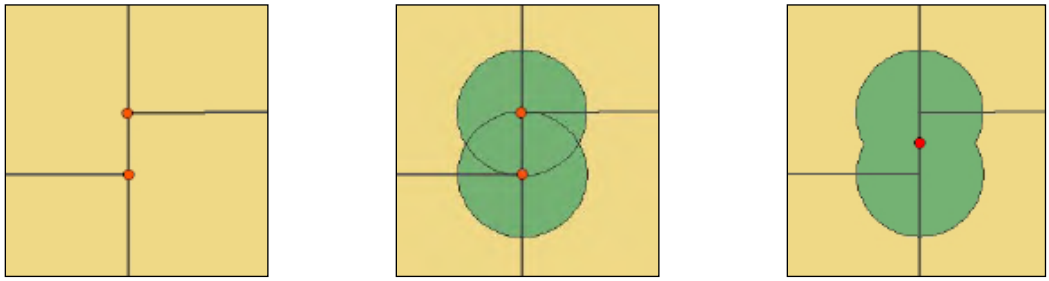
ZIP_HHENGL – Count of *English* speaking house-holds by linguistic isolation [P020002].
ZIP_HHSPAN – Count of *Spanish* speaking house-holds linguistic isolation [P020003].
ZIP_PCTLIN – Percent of population who are linguistically isolated [(P020004 + P020007 + P020010 + P020013) / P020001)].
ZIP_LINSPH – Count of house-holds that are linguistically isolated that speak *Spanish* [P020004].
ZIP_PSPLIN – Percent of house-holds that are linguistically isolated that speak *Spanish* [P020004] / [P020003] (total Spanish house-holds).
ZIP_LINEUR – Count of house-holds that are linguistically isolated that speak *Other Indo-European* languages [P020007].
ZIP_PIELIN – Percent of house-holds that are linguistically isolated that speak *Other Indo-European* languages [P020007] / [P020006] (total Other Indo-European house-holds).
ZIP_LINASN – Count of house-holds that are linguistically isolated that speak *Asian and Pacific Island* languages [P020010].
ZIP_PASLIN – Percent of house-holds that are linguistically isolated that speak *Asian and Pacific Island* languages [P020010] / [P020009] (total Asian and Pacific Island house-holds).
ZIP_LINOTR – Count of house-holds that are linguistically isolated that speak *Other* languages [P020013].
ZIP_POTLIN – Percent of house-holds that are linguistically isolated that speak *Other* languages [P020013] / [P020012] (total Other languages house-holds).

Built Environment Variables

Street Pattern Variables

ZIP_STRINT – Count of "unique" street intersections per ZIP Code 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.). The "unique" count takes into consideration that false positives can be created in regards to total intersections counts since judging an intersection can be complex, particularly in older areas where intersections are not regularly spaced, road rights of way have varying sizes, and centerlines may be slightly offset within the right of way. Therefore, I played with buffering each intersection with several different buffer sizes to help to deal with slightly offset street centerlines and divided roads (e.g., Broadway). Such situations result in multiple intersections where in reality they should be perceived as one intersection. After testing different sized buffered, I concluded that I needed to used 20-meter buffers around divided street nodes and 10-meter buffers around all other street nodes. After the buffers are created, they are merged where they overlap, boundaries are dissolved, and a centroid is created that represents the new street intersection. The figure below shows an example of where such methods would be used. This measure was calculated using the street centerline GIS layer from the New York State Accident Location Information System (NYS-ALIS). Prior to calculating this measure all primary highways with limited access and access ramps were removed since these street features are not used by pedestrians [ZIP_3WYINT + ZIP_4WYINT].

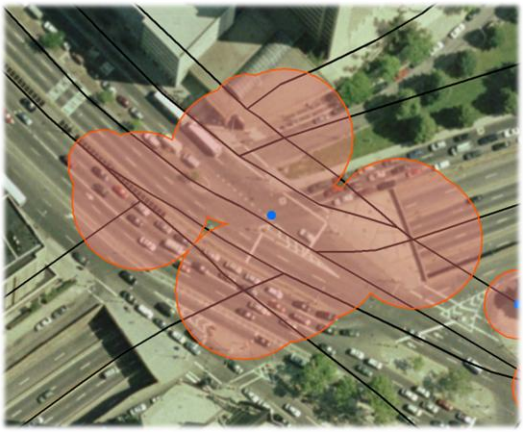
Figure. – Street intersection cleaning process.

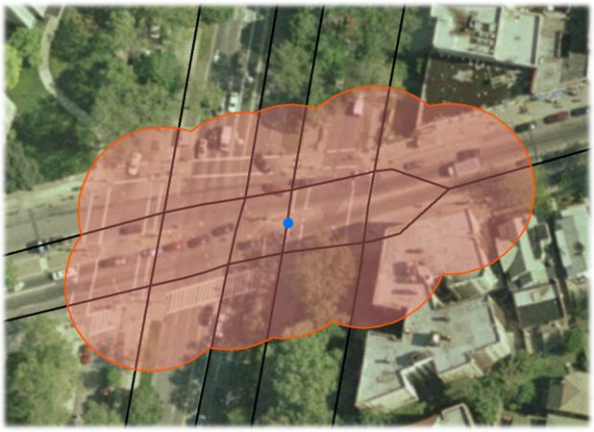


before buffering | after buffering by 10-meters | after merging, dissolving, and creating new centroid

ZIP_INTDEN – Density of "unique" streets intersections per ZIP Code in Km2 [ZIP_STRINT / LNDKM2].
ZIP_CULEND – Count of dead-ends and cul-de-sacs per ZIP Code [valence = 1].
ZIP_CULDEN – Density of dead-ends and cul-de-sacs per ZIP Code in Km2 [ZIP_CULEND / ZIP_LNDKM2].
ZIP_3WYINT – Count of three-way intersections per ZIP Code [valence = 3].
ZIP_3WYDEN – Density of three-way intersections per ZIP Code in Km2 [ZIP_3WYINT / ZIP_LNDKM2].
ZIP_3WYPCT – Percent of three-way intersections to all intersections = raw percentage of three-way intersections versus other intersections [valence 3+], with the added complication that some higher valence intersections are actually three-way and four-way intersections [ZIP_3WYINT / ZIP_STRINT].
ZIP_4WYINT – Count of four-way intersections per ZIP Code [valence = 4]. Four-way intersections are an indicator of grid street patterns, thought to be supportive for walking.
ZIP_4WYDEN – Density of four-way intersections per ZIP Code in Km2 [ZIP_4WYINT / ZIP_LNDKM2].
ZIP_4WYPCT – Percent of four-way intersections to all intersections = raw percentage of four-way intersections versus other intersections [valence 3+], with the added complication that some higher valence intersections are actually three-way and four-way intersections. Four-way intersections are an indicator of grid street patterns, believed by some to be more connected. One can imagine a high proportion of four-way intersections in an area, however, with enormous blocks, so the density of intersections should also be considered [ZIP_4WYINT / ZIP_STRINT].

PLEASE NOTE: As outlined in the discussion of the **ZIP_STRINT variable above, four-way and three-way intersection counts and percentages do not include intersections formed by primary highways with limited access and access ramps since these street features are not used by pedestrians, and counts reflect roads corrected for misaligned intersections by creating buffers around each intersection node, merging them where they overlapped, and creating a centroid for each new intersection polygon. See figures below for such examples.





ZIP_CNR – Connected node ratio is the number of street intersections (valence 3+) divided by the number of intersections plus cul-de-sacs (valence 1) (Dill 2003, 3) per ZIP Code. A connected node ratio close to 1.00 (the maximum value) means that there are few cul-de-sacs and, therefore, the street pattern is more highly connected. Dill (2003, 3) reports that ratios of 0.70 are preferred by those promoting highly-connected street patterns and recommends against networks with values less than 0.5. Of course there could be a perfect grid (CNR = 1.00(with enormous blocks that were hard to get around and such a pattern would fare well on this measure, so it is not a perfect indicator [ZIP_STRINT / [ZIP_CULEND + ZIP_STRINT]].

ZIP_STRKM – Total length of streets per ZIP Code in Km. Total street length was derived using the same shapefile described for the below ZIP_STRWID variable, using only those street centerline feature, where: FEATURE_CO = 2900 (Paved Road).
ZIP_STRDEN – Density of streets per ZIP Code in Km2 [ZIP_STRKM / ZIP_LNDKM2].
ZIP_STRKM2 – Total area of ZIP Code covered by street-network area (i.e., curb-to-curb roadbed) in Km2.
ZIP_STRPCT – Percent of total measurement geography covered by street-network area (i.e., curb-to-curb roadbed). The dark green areas in the figure to the right represents the total area of the street network while the orange are building footprint and the light green/tan is open area. To calculate the percentage street-network area, the street-network area layer was intersected with the different measurement geographies and the ratio of area occupied by street-network area calculated as a ratio [ZIP_STRKM2 / ZIP_LNDKM2].

ZIP_STRWID – Average street width per ZIP Code in Km. To determine the average street width the Streets Centerline layer was acquired from DoITT and all streets, where: FEATURE_CO = 2900 (Paved Road) where exported to a new shapefile. The newly exported shapefile was then intersected with the measurement geographies, roads dissolved by their unique and the average street width was derived using the available WIDTH variable (which is in feet but was converted to kilometers).

- DoITT GIS Web-Site: http://www.nyc.gov/html/doitt/html/eservices/eservices_gis_downloads.shtml
- Possible FEATURE_CO values:
  - 2245: Boardwalk
  - 2265: Interior Path
  - 2900: Paved Road
  - 2910: Unpaved Road
  - 2920: Alley
  - 2930: Stepped Street
  - 2240: Driveway

ZIP_STRMED – Median street width per ZIP Code in Km. The variable was calculated using the same DoITT Streets Centerline layer used to calculate the average street width variable.

Functional Road Classification Variables

Functional classifications of roads divide them into a hierarchy including subclasses of the road types described in detail following the listed variables. These measures assess the total length and ratio of the aforementioned functional classification road classes. Using the FCCs field from the road layers that were intersected with the network buffers, road segments of varying subclasses were summed and divided by the sum of the lengths of all roads to obtain the proportion of different subclass roads to all other roads.

ZIP_A1_KM – Total length in Km of **A1** – Primary Highways with Limited Access roads.
ZIP_A1_PCT – Total percentage of **A1** – Primary Highways with Limited Access roads.
ZIP_A2_KM – Total length in Km of **A2** – Primary Roads without Limited Access.
ZIP_A2_PCT – Total percentage of **A2** – Primary Roads without Limited Access.
ZIP_A3_KM – Total length in Km of **A3** – Secondary and Connecting Roads.
ZIP_A3_PCT – Total percentage of **A3** – Secondary and Connecting Roads.
ZIP_A4_KM – Total length in Km of **A4** – Local, Neighborhood, and Rural Roads.
ZIP_A4_PCT – Total percentage of **A4** – Local, Neighborhood, and Rural Roads.
ZIP_A5_KM – Total length in Km of **A5** – Vehicular Trails.
ZIP_A5_PCT – Total percentage of **A5** – Vehicular Trails.
ZIP_A6_KM – Total length in Km of **A6** – Roads with Special Characteristics.
ZIP_A6_PCT – Total percentage of **A6** – Roads with Special Characteristics.
ZIP_A7_KM – Total length in Km of **A7** – Roads as Other Thoroughfare.
ZIP_A7_PCT – Total percentage of **A7** – Roads as Other Thoroughfare.

- **A1 – Primary Highways With Limited Access—mostly interstates, include:**
  - **A10** Primary road with limited access, major category
  - **A11** Primary road with limited access or interstate hwy, unseparated
  - **A12** Primary road with limited access or interstate hwy, unseparated in tunnel
  - **A13** Primary road with limited access or interstate hwy, unseparated underpassing
  - **A14** Primary road with limited access or interstate hwy, unseparated rail line in center
  - **A15** Primary road with limited access or interstate hwy, separated
  - **A16** Primary road with limited access or interstate hwy, separated in tunnel
  - **A17** Primary road with limited access or interstate hwy, separated underpassing
  - **A18** Primary road with limited access or interstate hwy, separated rail line in center

- **A2 – Primary Roads Without Limited Access—mostly US and state highways, include:**
  - **A20** Primary Highways without limited access, major category
  - **A21** Primary Highways without limited access, unseparated
  - **A22** Primary Highways without limited access, unseparated in tunnel
  - **A23** Primary Highways without limited access, unseparated underpassing
  - **A24** Primary Highways without limited access, unseparated rail line in center
  - **A25** Primary Highways without limited access, separated
  - **A26** Primary Highways without limited access, separated in tunnel
  - **A27** Primary Highways without limited access, separated underpassing
  - **A28** Primary Highways without limited access, separated rail line in center

- **A3 – Secondary and Connecting Roads—mostly state and some county highways, include:**
  - **A30** Secondary state and county highways, major category
  - **A31** Secondary state and county highways, unseparated
  - **A32** Secondary state and county highways, unseparated in tunnel
  - **A33** Secondary state and county highways, unseparated underpassing
  - **A34** Secondary state and county highways, unseparated rail line in center
  - **A35** Secondary state and county highways, separated

A36 Secondary state and county highways, separated in tunnel
A37 Secondary state and county highways, separated underpassing
A38 Secondary state and county highways, separated rail line in center

- **A4 – Local, Neighborhood, and Rural Roads, include:**
  - **A40** Local, neighborhood, rural road, city-street, major category
  - **A41** Local, neighborhood, rural road, city-street, unseparated
  - **A42** Local, neighborhood, rural road, city-street, unseparated in tunnel
  - **A43** Local, neighborhood, rural road, city-street, unseparated underpassing
  - **A44** Local, neighborhood, rural road, city-street, unseparated rail line in center
  - **A45** Local, neighborhood, rural road, city-street, separated
  - **A46** Local, neighborhood, rural road, city-street, separated in tunnel
  - **A47** Local, neighborhood, rural road, city-street, separated underpassing
  - **A48** Local, neighborhood, rural road, city-street, separated rail line in center

- **A5 – Vehicular Trails, include:**
  - **A50** Vehicular (4WD) Trail, major category
  - **A51** Vehicular (4WD) Trail, unseparated
  - **A52** Vehicular (4WD) Trail, unseparated in tunnel
  - **A53** Vehicular (4WD) Trail, unseparated underpassing

- **A6 – Roads with Special Characteristics, include:**
  - **A60** At-grade ramp or connecting road not associated with a limited access highway
  - **A61** Cul-de-Sac, the closed end of a road that forms a loop or turn around
  - **A62** Traffic Circle, the portion of a road or intersection of roads forming a roundabout
  - **A63** Access Ramp, the portion of a road that forms a cloverleaf or limited access interchange
  - **A64** Service Road, provides access to businesses and rest areas
  - **A65** Ferry Crossing, Passenger, Seasonal
  - **A66** Ferry Crossing, Passenger, Year-Round
  - **A67** *not used*
  - **A68** Ferry Crossing, Vehicular, Seasonal
  - **A69** Ferry Crossing, Vehicular, Year-Round

- **A7 – Roads as Other Thoroughfare, include:**
  - **A70** Other Thoroughfare major category
  - **A71** Walkway, for pedestrians, usually unnamed
  - **A72** Stairway, stepped road for pedestrians, usually unnamed
  - **A73** Alley, road for service vehicles, located at the rear of buildings
  - **A74** Driveway, usually privately owned and unnamed
  - **A75** Road, parking area

Speed Limit Variable: Length-Weighted

MPHWGHTAVG – Length-weighted average speed limit per ZIP Code in miles per hour (mph).

$$WGHT\_SPEED = \frac{\sum_{j=1}^{N} LENGTH(S(j)) \times S(j)}{\sum_{j=1}^{N} LENGTH(S(j))}$$

Where: the N possible speed limits in the dataset are indexed 1,2,…,N (in this case N=7)

S = {1, 15, 25, 35, 45, 55, 65} is the set of possible speed limits (e.g. S(3) = 25)

LENGTH(S(j))= length in Km of roads with speed limit S(j) in buffer

Average Daily Traffic Variable: Length-Weighted

ADTWGHTAVG – Length-weighted average daily traffic per ZIP Code in number of vehicles.

Public Transportation Variables

♠ The New York City **Bus** information was originally obtained from the New York City Transit Authority in 1998. The data was realigned to the NYCMap basemap in 2002, and again updated in 2004.

BUS04_CNT – Count of unique MTA bus stops per ZIP Code as of 2004. This measure takes into account those bus stops that have multiple route opportunities at one location and each stop is counted only once regardless of how many route opportunities are available at any given bus stop.
BUS04_DEN – Density of unique bus stops per Km2 [BUS04_CNT / ZIP_LNDKM2].
BUS04_AVG – Kernel Density Estimation variable: **average** number of unique bus stops per Km2.
BUS04_MED – Kernel Density Estimation variable: **median** number of unique bus stops per Km2.

♠ 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.

SUBW07_CNT – Count of MTA subway stations per ZIP Code as of 2007.
SUBW07_DEN – Density of subway stations per Km2 [SUBW07_CNT /ZIP_LNDKM2].
SUBW07_AVG – Kernel Density Estimation variable: **average** number of subway stops per Km2.
SUBW07_MED – Kernel Density Estimation variable: **median** number of subway stops per Km2.

♠ Please note that the GIS Analyst is aware of the fact that the CHS data used for this project represents 2002-2006 and that the spatial dataset of NYC subway stations is from 2007. However, because subway station locations in NYC very, very rarely change, and a decision was made to use the 2007 subway station dataset because of its highly accurate map scale of 1:2,400. The accuracy of this dataset far exceeds the spatial precision of any other subway stations dataset in the BEH spatial data library.

Bicycle Route Variables: 2002 & 2007

New York City's bicycle network is a series of existing, proposed, and planned bicycle routes. These three categories can further be broken down into 1 of 6 route categories:

- **Class 1 Routes**: On-Street Striped - Part of the roadway and delineated by pavement markings and regulatory signage. The lane which can be shared with in-line skaters, is usually located next to curb lane parking, and may include a marked buffer zone.
- **Class 2 Routes**: On-Street Proposed Route - A proposed on-street bicycle route (no pavement markings) that is recommended for bicycle travel. These routes may become On-Street Striped routes at which point their classification will change.
- **Class 3 Routes**: Greenway Paths - Separated from the roadway and delineated by pavement markings and regulatory signage. Bicycle paths are usually shared with multiple users, including pedestrians, runners and skaters. These paths usually follow waterfronts, rail lines, highways and parks.

- **Class 4 Routes**: Greenway Connectors/Signed Route - Shared use of the roadway, typically designated with informational signs. These routes often connect to greenway paths, but may also serve as a connection between on-street bike lanes.
- **Class 5 Routes**: Planned/Proposed Greenways - A proposed bicycle path separated from the roadway, these paths usually follow waterfronts, rail lines, highways and parks. These routes may become Greenway Paths in the future at which point their classification will change.
- **Class 6 Routes**: Signed Route - Shared use of the roadway, typically designated with informational signs.

Using the original PDF bicycle route maps of NYC from 2002, bicycle routes were digitized and attributed appropriately using the above 6 route categories. Bicycle routes from 2007 were acquired in shapefile format from the NYD Department of City Planning and were already digitized and attributed with the above 6 route categories.

♠ Variables are available for 2002 and 2007, where "YY" in the variable name denotes the year.

Existing Bicycle Route Variables

BIKLENYYC1 – Total length in Km of **Class 1** bicycle routes per ZIP Code.
BIKPCTYYC1 – Total percentage of **Class 1** bicycle routes [BIKLENYYC1 / BIKLENYY_T].
BIKLENYYC3 – Total length in Km of **Class 3** bicycle routes per ZIP Code.
BIKPCTYYC3 – Total percentage of **Class 3** bicycle routes [BIKLENYYC3 / BIKLENYY_T].
BIKLENYYC4 – Total length in Km of **Class 4** bicycle routes per ZIP Code.
BIKPCTYYC4 – Total percentage of **Class 4** bicycle routes [BIKLENYYC4 / BIKLENYY_T].
BIKLENYYC6 – Total length in Km of **Class 6** bicycle routes per ZIP Code.
BIKPCTYYC6 – Total percentage of **Class 6** bicycle routes [BIKLENYYC6 / BIKLENYY_T].
BIKLENYY_A – Total length in Km of **Existing** bicycle routes per ZIP Code.
BIKPCTYY_A – Total percentage of **Existing** bicycle routes [BIKLENYY_A / BIKLENYY_T].
BIKDENYY_A – Density of **Existing** bicycle routes per Km2 [BIKLENYY_A / ZIP_LNDKM2].

Proposed or Planned Bicycle Route Variables

BIKLENYYC2 – Total length in Km of **Class 2** bicycle routes per ZIP Code.
BIKPCTYYC2 – Total percentage of **Class 2** bicycle routes [BIKLENYYC2 / BIKLENYY_A].
BIKLENYYC5 – Total length in Km of **Class 5** bicycle routes per ZIP Code.
BIKPCTYYC5 – Total percentage of **Class 5** bicycle routes [BIKLENYYC5 / BIKLENYY_A].
BIKLENYY_P – Total length in Km of **Proposed** or **planned** bicycle routes per ZIP Code.
BIKPCTYY_P – Total percentage of **Proposed** or **Planned** bicycle routes [BIKLENYY_P / BIKLENYY_T].
BIKDENYY_P – Density of **Proposed** or **Planned** bicycle routes per Km2 [BIKLENYY_P / ZIP_LNDKM2].

Existing, Proposed, or Planned Bicycle Route Variables

BIKLENYY_T – Total length in Km of **Existing**, **Proposed**, or **Planned** bicycle routes per ZIP Code.
BIKDENYY_T – Density of **Existing**, **Proposed**, or **Planned** bicycle routes per Km2 [BIKLENYY_T / ZIP_LNDZM2].

Dun and Bradstreet Variables: 2001 & 2005

Dun & Bradstreet <www.dnb.com>, a commercial vendor of business data, maintains a comprehensive database of micro-data on over 11 million U.S. business locations. This data was purchased for NYC and represents the food and retail presence on the ground. Variables are available for 2001 and 2005, where "Y" in the variable name denotes the last digit of the respective year.

Food Environment Categories: Point-in-Polygons Counts

DUNBRADTOY – Count of Dun & Bradstreet businesses per ZIP Code. Included SIC codes at the 4-digit level: **5411** – Grocery Stores • **5421** – Meat and Fish Markets • **5431** – Fruit and Vegetable Markets • **5441** – Candy, Nut, and Confectionery Store • **5451** – Dairy Products Stores • **5461** – Retail Bakeries • **5499** – Miscellaneous Food Stores • **5812** – Eating Places.

"PRIM_SIC" = '5411' OR "PRIM_SIC" = '5421' OR "PRIM_SIC" = '5431' OR "PRIM_SIC" = '5441' OR "PRIM_SIC" = '5451' OR "PRIM_SIC" = '5461' OR "PRIM_SIC" = '5499' OR "PRIM_SIC" = '5812'

EATPLC_TOY – Count of Dun & Bradstreet eating place businesses [SIC = 5812] per ZIP Code. "PRIM_SIC" = '5812'

EATPLCXTRY – EATPLC_TOT plus national chains, local chains, or pizza (selected by name) records that do not have the eating place SIC code 5812 (e.g. a Popeye's that classified itself as a drinking place).
"AFASTPIZZA" = 1 OR "PRIM_SIC" = '5812'

OTHERRESY – Restaurants NOT classified in one of the following fast food or pizza categories ("EATPLC_TOT" =1 and removed records where: "NATL_CHAIN" =1 OR "LOCFASTCHN" =1 OR "LOCSICFAST" =1 OR "LOCFASTFOD" =1 OR "ALLFASTFOD" =1 OR "PIZZA" =1 OR "NTLCHNPIZA" =1 OR "AFASTPIZZA" =1). These are all restaurants not classified as fast food by any of our possible definitions.

NATL_CHAIY: National chain restaurant variable [If TRADESTYLE or COMPNAME = ARBY'S, AU BON PAIN, BASKIN-ROBBINS, BEN & JERRY'S, BLIMPIE SUBS & SALADS, BOJANGLES, BOSTON MARKET, BURGER KING, CARVEL ICE CREAM CAKES, CHECKERS, CHIPOTLE, CHURCHS CHICKEN, COLD STONE CREAMERY, COSI, DAIRY QUEEN, DAYLIGHT DONUTS, DOMINO'S PIZZA, DUNKIN' DONUTS, EL POLLO LOCO, HAAGEN-DAZS, HARDEE'S, I CAN'T BELIEVE IT'S YOGURT, JAMBA JUICE, KFC, KRISPY KREME, LITTLE CAESARS PIZZA, LONG JOHN SILVER'S, MC DONALD'S, PANDA EXPRESS, PAPA JOHN'S PIZZA, PIZZA DELIGHT, PIZZA HUT, POPEYE'S, QUIZNOS, RED ROBIN, ROUND TABLE PIZZA, SBARRO, SCHLOTZSKY'S DELI, STARBUCKS, SUBWAY, TACO BELL, TCBY, TOGOS, WENDY'S, WHITE CASTLE].

This list is a combination of Catie's coding and the franchises included in InfoUSA's franchise codes. Only a few were added from the InfoUSA list because Catie's list had the majority.

See SPSS code for selection criteria (Code for Fast Food Variables). After applying the code, went through and removed the records where the SIC code was not an "eating place" or one of the other food SIC codes.
Selected NATL_CHAIN = 1, selected from selection the following:
"PRIM_SIC" = '5411' OR "PRIM_SIC" = '5421' OR "PRIM_SIC" = '5431' OR "PRIM_SIC" = '5441' OR "PRIM_SIC" = '5451' OR "PRIM_SIC" = '5461' OR "PRIM_SIC" = '5499' OR "PRIM_SIC" = '5812'
Switched the selection and then selected NATL_CHAIN=1 from this switched selection – marked these '0'. Did the same for LOCFASTCHN.

LOCFASTCHY: This variable identifies local fast food by name. NATL_CHAIN = 0 AND [IF TRADESTYLE OR COMP_NAME = Crown Fried Chicken, Kennedy Fried Chicken, Kosher Delight Corp, Miami Subs & Grill, El Pollo Supremo, Supreme Chicken Of New Jersey, Nathans Famous Inc, Pudgies Famous Chicken, Zorn Famous Chicken & Ribs, Grays Papaya, Manhattan Bagel, Roy Rogers, Everything Yogurt & Salad, Coffee Shop North Central Hosp, Directors Metro Food Service, Chicken Holiday, Auntie Annes, Metropolitan Deli, Xando, OR Papaya King].

LOCSICFASY: This variable identifies local fast food by SIC code—restaurants that self identified as fast food via their SIC coding.
By SIC coding: [IF (NATL_CHAIN = 0) AND SIC = 58120300, 58120307, 58120308, 58120302]
Query code: "PRIM8_EXT1" ='58120300' OR "PRIM8_EXT2" ='58120300' OR "PRIM8_EXT3" ='58120300' OR "PRIM8_EXT4" ='58120300' OR "PRIM8_EXT1" ='58120307' OR "PRIM8_EXT2" ='58120307' OR "PRIM8_EXT3" ='58120307' OR "PRIM8_EXT4" ='58120307' OR "PRIM8_EXT1" = '58120308' OR "PRIM8_EXT2" ='58120308' OR "PRIM8_EXT3" ='58120308' OR "PRIM8_EXT4" ='58120308' OR "PRIM8_EXT1" = ' 58120302' OR "PRIM8_EXT2" =' 58120302' OR "PRIM8_EXT3" =' 58120302' OR "PRIM8_EXT4" =' 58120302'

LOCFASTFOY: The combination of LOCFASTCHN and LOCSICFAST. Restaurants that were either a local fast food chain (by name) or self-identified as fast food. "LOCSICFAST" =1 OR "LOCFASTCHN" =1

ALLFASTFOY – Variable that puts national and local fast food into one variable.
"NATL_CHAIN" =1 OR "LOCFASTFOD" =1

PIZZAY – Pizza variable [IF (NATL_CHAIN = 0 and LOCFASTFOD=0) AND SIC = 58120600, 58120601, 58120602, OR TRADESTYLE = PIZZA OR COMP_NAME = PIZZA].

Query code: "PRIM8_EXT1" ='58120600' OR "PRIM8_EXT2" ='58120600' OR "PRIM8_EXT3" ='58120600' OR "PRIM8_EXT4" = '58120600' OR "PRIM8_EXT1" ='58120601' OR "PRIM8_EXT2" ='58120601' OR "PRIM8_EXT3" ='58120601' OR "PRIM8_EXT4" ='58120601' OR "PRIM8_EXT1" ='58120602' OR "PRIM8_EXT2" ='58120602' OR "PRIM8_EXT3" ='58120602' OR "PRIM8_EXT4" ='58120602' OR "COMP_NAME" LIKE '%PIZZA%' OR "TRADESTYLE" LIKE '%PIZZA%'

NTLCHNPIZY – National chain and pizza variable [OF ((NATL_CHAIN = 1) OR (PIZZA = 1))].

"NATL_CHAIN" =1 OR "PIZZA" =1

AFASTPIZZY – All fast food and pizza places [IF (ALLFASTFOD = 1 OR PIZZA = 1)].

"ALLFASTFOD" =1 OR "PIZZA" =1

SUPRMARKEY: All supermarkets where SIC = 5411, sales volume is >= $2,000,000 or total number of employees is >= 18.

Query code: "PRIM_SIC" ='5411' AND "SALES_VOL" >='000000002000000' –marked these.
"PRIM_SIC" = '5411' AND "SALES_VOL" = '000000000000000' AND "EMPL_HERE" >= '000000018' – marked these.

GROCERYY: All grocery stores where SIC = 5411, sales volume is < $2,000,000 and the total number of employees is < 18 and > 4. The selection is less than $2 mil AND between 17 and 5 employees because if you choose $2 mil OR between 17 and 5 employees, you will also capture the < 5 employee establishments, which is how bodegas have been defined.

Query code: "PRIM_SIC" ='5411' AND "SALES_VOL" < '000000002000000' AND "EMPL_HERE" < '000000018' AND "EMPL_HERE" > '000000004'.
- Also had to mark stores as grocery stores when their sales volumes were less than $2 mil (and not = $0) and their number of employees was equal to or greater than 18.
- "PRIM_SIC" = '5411' AND "SALES_VOL" < '000000002000000' AND "EMPL_HERE" >= '000000018' AND "SALES_VOL" <> '000000000000000'

BODEGASY: All bodegas where SIC = 5411 and the total number of employees is <= 4 and Grocery does not equal 1 and supermarket does not equal 1.

Query code: "PRIM_SIC" ='5411' AND "EMPL_HERE" <= '000000004' AND "SUPRMARKET" <>1 AND "GROCERY" <>1.

CONVENIENY: All conveniences stores where SIC = 541102 (in any of the 4 SIC fields). Convenience stores are mutually exclusive with supermarkets, grocery stores, and bodegas. Removed convenience stores from supermarkets, grocery, and bodegas, so places that code themselves as convenience stores are first convenience stores.

Query code: "PRIM8_EXT1" LIKE '541102%' OR "PRIM8_EXT2" LIKE '541102%' OR "PRIM8_EXT3" LIKE '541102%' OR "PRIM8_EXT4" LIKE '541102%'

FRUITVEGIY: Fruit and vegetable markets where SIC = 5431 ("PRIM_SIC" ='5431').

FISHY – Fish/seafood markets where SIC6_EXT1 = 542101 OR SIC6_EXT2 = 542101 OR SIC6_EXT3 = 542101 OR SIC6_EXT4 = 542101.

"PRIM8_EXT1" LIKE '542101%' OR "PRIM8_EXT2" LIKE '542101%' OR "PRIM8_EXT3" LIKE '542101%' OR "PRIM8_EXT4" LIKE '542101%'

MEATY – Meat markets where SIC6_EXT1 = 542102 OR SIC6_EXT2 = 542102 OR SIC6_EXT3 = 542102 OR SIC6_EXT4 = 542102.

"PRIM8_EXT1" LIKE '542102%' OR "PRIM8_EXT2" LIKE '542102%' OR "PRIM8_EXT3" LIKE '542102%' OR "PRIM8_EXT4" LIKE '542102%'

MEATORFISY: Meat or fish markets that are not classified beyond the 4-digit primary SIC code in addition to those that are (either Fish or Meat markets from above).

"PRIM_SIC" ='5421' AND "FISH" = 1 OR "MEAT" = 1

CANDYNUTY – Candy, nut, and confectionery stores where SIC = 5441.
"PRIM_SIC" ='5441'

BAKERYY: Retail bakeries where SIC = 5461.
"PRIM_SIC" ='5461' AND "NATL_CHAIN" <>1
also removed Pizza and Local Chain fast food (17 records) from the bakery category

NATURALFOY: Natural food stores including health and dietetic food stores [SIC = 549901].
"PRIM8_EXT1" LIKE '549901%' OR "PRIM8_EXT2" LIKE '549901%' OR "PRIM8_EXT3" LIKE '549901%' OR "PRIM8_EXT4" LIKE '549901%'

SPECIALTYY: Special stores including dairy products stores [SIC = 5451] and miscellaneous food stores [SIC = 5499] excluding health and dietetic food stores [SIC = 549901] and fast food or pizza.
"PRIM_SIC" ='5451' OR "PRIM_SIC" ='5499' AND "NATURALFOD" <>1
removed from the selection: "NATL_CHAIN" =1 OR "LOCFASTCHN" =1 OR "PIZZA" =1

DRINKINGY – Drinking places where SIC = 5813
"PRIM_SIC" ='5813'

Checked each variable and fixed overlaps using the following logic rules:

- National Chain or Pizza → National Chain
- Local fast food (either by name or SIC) or Pizza → Local fast food
- National Chain or Bakery → National Chain
- Local Chain or Bakery → Local Chain
- Pizza or Bakery → Pizza
- Bodega or Pizza → Bodega (same goes for other grocery stores or convenience stores)

HEALTHYY – Count of BMI-healthy food outlets (NYC Wide–2001: n=1,514; 2005: n=1,833). This category includes Supermarkets, Fruit and Veggie Markets, and Natural Food Stores ["SUPRMARKET" = 1 OR "FRUITVEGIE" = 1 OR "NATURALFOD" = 1].

UNHEALTHYY – Count of BMI-unhealthy food outlets (NYC Wide–2001: n=12,071; 2005: n=17,348). This category includes Fast Food Restaurants, Pizza Restaurants, Convenience Stores, Bodegas, Bakeries, Candy and Nut Stores, and Meat Markets ["BAKERY" = 1 OR "CANDYNUT" = 1 OR "MEAT" = 1 OR "CONVENIENC" = 1 OR "BODEGAS" = 1 OR "AFASTPIZZA" = 1].

NEUTRALY – Count of BMI-neutral or intermediate food outlets (NYC Wide–2001: n=15,780; 2005: n=20,671). This category includes Other Restaurants (non-Fast Food), Grocery Stores, Fish Markets, and Specialty Food Stores ["GROCERY" = 1 OR "FISH" = 1 OR "SPECIALTY" = 1 OR "OTHERREST" = 1].

Food Environment Categories: Kernel Density Measures

DBYY_H_AVG – Kernel Density Estimation variable: **average** number of BMI-healthy food outlets per Km2.
DBYY_H_MED – Kernel Density Estimation variable: **median** number of BMI-healthy food outlets per Km2.
DBYY_N_AVG – Kernel Density Estimation variable: **average** number of BMI-neutral food outlets per Km2.
DBYY_N_MED – Kernel Density Estimation variable: **median** number of BMI-neutral food outlets per Km2.
DBYY_U_AVG – Kernel Density Estimation variable: **average** number of BMI-unhealthy food outlets per Km2.
DBYY_U_MED – Kernel Density Estimation variable: **median** number of BMI-unhealthy food outlets per Km2.

Retail Environment Categories: Point-in-Polygons Counts

HARDWAREX – Hardware stores where SIC = 5251.
"PRIM_SIC" ='5251'

DRUG_STORX – Drug stores and proprietary stores where SIC = 5912.
"PRIM_SIC" ='5912'

BANKX – National commercial banks, state commercial banks, commercial banks (nec), federal savings institutions, and savings institutions (except federal) where SIC = 6021 OR SIC = 6022 OR SIC = 6029 OR SIC = 6035 OR SIC = 6036.
"PRIM_SIC" ='6021' OR "PRIM_SIC" ='6022' OR "PRIM_SIC" ='6029' OR "PRIM_SIC" ='6035' OR "PRIM_SIC" ='6036'

CREDIT_UX – Federal and state credit unions where SIC = 6061 OR SIC = 6062.
"PRIM_SIC" ='6061' OR "PRIM_SIC" ='6062'

CHILDCAREX – Child day care facilities where SIC = 8351.
"PRIM_SIC" ='8351'

LAUNDRYX – Coin operated laundries and cleaning facilities where SIC = 7215.
"PRIM_SIC" ='7215'
GROOMINGX – Beauty and barber shops where SIC = 7231 OR SIC = 7241.
"PRIM_SIC" ='7231' OR "PRIM_SIC" ='7241'

Sidewalk Café Variables

Sidewalk cafés were geocoded from a list of Legally Operating New York City Sidewalk Cafés (as of June 8, 2006) which was provided to BEH by the New York City Department of Consumer Affairs. For more information on Legally Operating Sidewalk Cafés, please visit http://home2.nyc.gov/html/dca/html/licenses/swc_main.shtml

CAFE06_CNT – Count of sidewalk cafés per ZIP Code as of 2006.
CAFE06_DEN – Density of sidewalk cafés per Km2 [CAFE06_CNT / ZIP_LNDKM2].
CAFE06_AVG – Kernel Density Estimation variable: **average** number of sidewalk cafés per Km2.
CAFE06_MED – Kernel Density Estimation variable: **median** number of sidewalk cafés per Km2.

Poultry and Small Animal Slaughterhouse Variables

A list of licensed Poultry and Small Animal Slaughterhouses were acquired from Jennifer Jensen (j_jensen@planning.nyc.gov) with the Division of Housing, Economic & Infrastructure Planning, Department of City Planning representing Slaughterhouses as of 2004 (a more specific date in 2004 is not known). This data was then cleaned and geocoded using GeoSupport software.

MEAT04_CNT – Count of slaughterhouses per ZIP Code as of 2004.
MEAT04_DEN – Density of slaughterhouses per Km2 [MEAT04_CNT / ZIP_LNDKM2].
MEAT04_AVG – Kernel Density Estimation variable: **average** number of slaughterhouses per Km2.
MEAT04_MED – Kernel Density Estimation variable: **median** number of slaughterhouses per Km2.

Farmers' Market Variables

Lists of Farmers' Markets were acquired and reconciled for 2006 from various sources, including:

- New York City Coalition Against Hunger (http://www.nyccah.org/)
- Council on the Environment of New York City (http://www.cenyc.org/)
- Farmers' Market Federation of New York (http://www.nyfarmersmarket.com/)

FMRK06_CNT – Count of farmers' markets per ZIP Code as of 2006.
FMRK06_DEN – Density of farmers' markets per Km2 [FMRK06_CNT / ZIP_LNDKM2].
FMRK06_AVG – Kernel Density Estimation variable: **average** number of farmers markets per Km2.
FMRK06_MED – Kernel Density Estimation variable: **median** number of farmers markets per Km2.

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.

ZIP_RESKM2 – Residential Building area in Km2 (ZIP_LUMIXA input variable).
ZIP_RESPCT – % Residential Building area [([ZIP_RESKM2] / ([ZIP_COMKM2] + [ZIP_RESKM2])) * 100].
ZIP_COMKM2 – Commercial Building area in Km2 (ZIP_LUMIXA input variable).
ZIP_COMPCT – % Commercial Building area [([ZIP_COMKM2] / ([ZIP_RESKM2] + [ZIP_COMKM2])) * 100].
ZIP_LUMIXA – Land Use Mix A. This measure was calculated using the PLUTO tax lot data available from the NYC Department of City Planning. A co-distribution of commercial and residential building area was derived from the PLUTO 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.

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

ZIP_LU01KM – One and Two Family Building area Km2.
ZIP_LU01PT – Percent of One and Two Family Building area [ZIP_LU01KM / ZIP_LNDKM2].
ZIP_LU02KM – Multi–Family Walk-up Building area in Km2.
ZIP_LU02PT – Percent of Multi–Family Walk-up Building area [ZIP_LU02KM / ZIP_LNDKM2].
ZIP_LU03KM – Multi–Family Elevator Building area in Km2.
ZIP_LU03PT – Percent of Multi–Family Elevator Building area [ZIP_LU03KM / ZIP_LNDKM2].
ZIP_LU04KM – Mixed Residential and Commercial Building area in Km2.
ZIP_LU04PT – Percent of Mixed Residential and Commercial Building area [ZIP_LU04KM / ZIP_LNDKM2].
ZIP_LU05KM – Commercial and Office Building area in Km2.
ZIP_LU05PT – Percent of Commercial and Office Building area [ZIP_LU05KM / ZIP_LNDKM2].
ZIP_LU06KM – Industrial and Manufacturing area in Km2.
ZIP_LU06PT – Percent of Industrial and Manufacturing area [ZIP_LU06KM / ZIP_LNDKM2].
ZIP_LU07KM – Transportation and Utility area in Km2.
ZIP_LU07PT – Percent of Transportation and Utility area [ZIP_LU07KM / ZIP_LNDKM2].

ZIP_LU08KM – Public Facilities and Institutions area in Km2.
ZIP_LU08PT – Percent of Public Facilities and Institutions area [ZIP_LU08KM / ZIP_LNDKM2].
ZIP_LU09KM – Open Space and Outdoor Recreation area in Km2.
ZIP_LU09PT – Percent of Open Space and Outdoor Recreation area [ZIP_LU09KM / ZIP_LNDKM2].
ZIP_LU10KM – Parking Facilities area in Km2.
ZIP_LU10PT – Percent of Parking Facilities area [ZIP_LU10KM / ZIP_LNDKM2].
ZIP_LU11KM – Vacant Land area in Km2.
ZIP_LU11PT – Percent of Vacant Land area [ZIP_LU11KM / ZIP_LNDKM2].
ZIP_LUORKM – Other area in Km2 (parcels not classified as category 1-11 but still a legal tax lot).
ZIP_LUORPT – Percent of Other area [ZIP_LUOTR / ZIP_LNDKM2].
ZIP_LUTOT – Total area of all 11 categories plus the Other Land Use category in Km2.
ZIP_LUPTOT – Percent of the 11 categories plus the Other Land Use category area [ZIP_LUTOT / ZIP_LNDKM2].
ZIP_ROWKM2 – Total Right-of-Way area in Km2 (areas not defined as tax lots, e.g., streets, sidewalks, etc.),
ZIP_ROWPCT – Percent Right-of-Way area [ZIP_AREAROW / ZIP_LNDKM2].

Street Tree Variables: 2005-2006

Street trees are thought to improve pedestrian comfort (shade, wind protection) and the attractiveness of the street. These are measures of the total number of street trees, density of street trees by ZIP Code area in Km2, density of street trees by total length of all streets per ZIP Code in Km, and the average and median number of street trees per ZIP Code per Km2 as derived from kernel density grid surfaces.

Point-level data from the City of New York, Parks & Recreation's 2005-06 Street Tree Census was used to create the metrics. To learn more, please visit: http://www.nycgovparks.org/sub_your_park/trees_greenstreets/treescount/.

TREE05_CNT – Count of street trees per ZIP Code as of 2005-06.
TREE05_DEN – Density of street trees per Km2 of land area [TREE05_CNT / ZIP_LNDKM2].
TREE05_STR – Density of street trees per Km of road length [TREE05_CNT / ZIP_STRKM].
TREE05_AVG – Kernel Density Estimation variable: **average** number of street trees per Km2.
TREE05_MED – Kernel Density Estimation variable: **median** number of street trees per Km2.

Homicide Crime Variables: 2003-2006

The metrics available here are of the total number of homicides, density of homicides by ZIP Code area in Km2, and the average and median number of homicides per ZIP Code per Km2 as derived from kernel density grid surfaces for individual years and averaged across all years.

The locations of these homicide variables were mined from this New York Times web-site: http://projects.nytimes.com/crime/homicides/map/

HOMC03_CNT – Count of homicides in 2003 per ZIP Code.
HOMC03_DEN – Density of homicides per Km2 [HOMC03_CNT / ZIP_LNDKM2].
HOMC03_AVG – Kernel Density Estimation variable: **average** number of homicides per Km2.
HOMC03_MED – Kernel Density Estimation variable: **median** number of homicides per Km2.

HOMC04_CNT – Count of homicides in 2004 per ZIP Code.
HOMC04_DEN – Density of homicides per Km2 [HOMC04_CNT / ZIP_LNDKM2].
HOMC04_AVG – Kernel Density Estimation variable: **average** number of homicides per Km2.
HOMC04_MED – Kernel Density Estimation variable: **median** number of homicides per Km2.

HOMC05_CNT – Count of homicides in 2005 per ZIP Code.
HOMC05_DEN – Density of homicides per Km2 [HOMC05_CNT / ZIP_LNDKM2].
HOMC05_AVG – Kernel Density Estimation variable: **average** number of homicides per Km2.
HOMC05_MED – Kernel Density Estimation variable: **median** number of homicides per Km2.

HOMC06_CNT – Count of homicides in 2006 per ZIP Code.
HOMC06_DEN – Density of homicides per Km2 [HOMC06_CNT / ZIP_LNDKM2].
HOMC06_AVG – Kernel Density Estimation variable: **average** number of homicides per Km2.
HOMC06_MED – Kernel Density Estimation variable: **median** number of homicides per Km2.

HOMCIDECNT – Count of homicides in 2003-06 per ZIP Code [HOMC03_CNT + HOMC04_CNT + HOMC05_CNT + HOMC06_CNT].
HOMCIDEDEN – Density of homicides per Km2 [(HOMC03_CNT + HOMC04_CNT + HOMC05_CNT + HOMC06_CNT) / ZIP_LNDKM2].
HOMCIDEAVG – Kernel Density Estimation variable: **average** number of homicides per Km2 [(HOMC03_AVG + HOMC04_AVG + HOMC05_AVG + HOMC06_AVG) / 4].
HOMCIDEMED – Kernel Density Estimation variable: **median** number of homicides per Km2.

Pedestrian–Bicyclist–Motorist Crash Variables

Crash data comes from the New York State Department of Transportation. Variables are aggregated by both pedestrian—motorist injuries, pedestrian—motorist fatalities, bicyclist—motorist injuries, and bicyclist—motorist fatalities. Variables are available for 2002 through 2006, where "YY" in the variable name denotes the year.

PINJYY_CNT – Count of pedestrian-motorist injuries per ZIP Code.
PINJYY_DEN – Density of pedestrian-motorist injuries per Km2 [PINJYY_CNT / ZIP_LNDKM2].
BINJYY_CNT – Count of bicyclist-motorist injuries per ZIP Code.
BINJYY_DEN – Density of bicyclist-motorist injuries per Km2 [BINJYY_CNT / ZIP_LNDKM2].
PKILYY_CNT – Count of pedestrian-motorist fatalities per ZIP Code.
PKILYY_DEN – Density of pedestrian-motorist fatalities per Km2 [PKILYY_CNT / ZIP_LNDKM2].
BKILYY_CNT – Count of bicyclist-motorist fatalities per ZIP Code.
BKILYY_DEN – Density of bicyclist-motorist fatalities per Km2 [BKILYY_CNT / ZIP_LNDKM2].

Department of Sanitation Sidewalk and Street Cleanliness Scorecard Variables

Project Scorecard Overview

Since 1973, the City of New York has used a litter measurement tool called "Project Scorecard" to evaluate and measure the cleanliness of its streets and sidewalks. Trained evaluation teams use the Scorecard to "rate" the degree of surface litter by comparing actual street conditions to photographic standards. This measurement tool provides objective information about street and sidewalk cleanliness. Project Scorecard is managed by the Mayor's Office of Operations because they are responsible for tracking and monitoring the City's cleanliness over time. To learn more about the purpose, measurement methods, etc. of the Project Scorecard, please visit: <http://www.usmayors.org/USCM/best_practices/litter/NewYork.html>.

Sidewalk and Street Scorecard Variables: Individual Years

Scorecard variables are aggregated to 1 of 234 NYC Sanitation Sections. Variables are available from 2002-2006, where "YY" in the variable name denotes the year. These variables were created using an area-weighted methodology as Sanitation Sections do not nest within ZIP Codes.

STCLNYYAVG – Mean street cleanliness as a numerical value ranging from 1.0 [cleanest] to 3.0 [filthiest]. Ratings below 1.5 are considered "acceptable" according to both the consensus arrived at in public surveys conducted by Scorecard and the Department's own operating standards.

STACCYYPCT – Percent streets acceptable as a range of 0.0 to 100.0.
STFILYYPCT – Percent streets filthy as a range of 0.0 to 100.0.
SWCLNYYAVG – Mean sidewalk cleanliness as a numerical value ranging from 1.0 [cleanest] to 3.0 [filthiest]. Ratings below 1.5 are considered "acceptable" according to both the consensus arrived at in public surveys conducted by Scorecard and the Department's own operating standards.
SWACCYYPCT – Percent sidewalks acceptable as a range of 0.0 to 100.0.
SWFILYYPCT – Percent sidewalks filthy as a range of 0.0 to 100.0.

Sidewalk and Street Scorecard Variables: Averaged Across Years

Scorecard variables were also averaged across 2002-2006.

STCLEANWGT – Mean street cleanliness as a numerical value ranging from 1.0 [cleanest] to 3.0 [filthiest]. Ratings below 1.5 are considered "acceptable" according to both the consensus arrived at in public surveys conducted by Scorecard and the Department's own operating standards.
STACCPTWGT – Percent streets acceptable as a range of 0.0 to 100.0.
STFILTYWGT – Percent streets filthy as a range of 0.0 to 100.0.
SWCLEANWGT – Mean sidewalk cleanliness as a numerical value ranging from 1.0 [cleanest] to 3.0 [filthiest]. Ratings below 1.5 are considered "acceptable" according to both the consensus arrived at in public surveys conducted by Scorecard and the Department's own operating standards.
SWACCPTWGT – Percent sidewalks acceptable as a range of 0.0 to 100.0.
SWFILTYWGT – Percent sidewalks filthy as a range of 0.0 to 100.0.

Walkability Index Variables



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.

To date, BEH has created and used two different versions of the "Walkability Index." These indexes were created by z-scoring and then summing and/or averaging various components of the built and social environments which are further discussed below. Mike Bader will provide a Stata .do file that documents the components used for each version and also performs the z-scoring and performs the construction of the two "Walkability Index" variables.

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.

Several of the variables employed in the "Walkability Index" were previously calculated and are aforementioned in this data dictionary. To avoid duplication of variables in the dataset itself, the "Walkability Index" variables flagged with a "♦" are also found in other sections of this data dictionary but are listed again here in an abridged definition for the sake of convenience. The variables are not, however, duplicated in the dataset.

♦ **ZIP_LNDKM2** – Total land area in Km2 (inland water features subtracted out).
♦ **ZIP_POPDEN** – Population density per Km2.
♦ **ZIP_STRINT** – Count of unique street intersections.

♦ **ZIP_INTDEN** – Density of unique streets intersections per Km2.
♦ **BUS04_CNT** – Count of unique MTA bus stops.
♦ **BUS04_DEN** – Density of unique bus stops per Km2.
♦ **SUBW07_CNT** – Count of MTA subway stations.
♦ **SUBW07_DEN** – Density of subway stations per Km2.
♦ **ZIP_COMPCT** – Percent Commercial Building area.
♦ **ZIP_LUMIXA** – Land Use Mix A. This measure was calculated using the PLUTO tax lot data available from the NYC Department of City Planning. A co-distribution of commercial and residential building area was derived from the PLUTO 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.

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

RETLFAR – Retail floor area ratio – Retail building floor area divided by retail land area, both measures drawn from the MapPLUTO database.
RESDEN1 – Density of residential units per residential building area – the number of residential units over the total residential building floor area measurement geography in Km2 [RES_UNITS / B3].
RESDEN2 – Density of residential units – the number of housing units divided by the total ZIP Code land area in Km2 [RES_UNITS / ZIP_LNDKM2].
ENTROPY – 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 PLUTO database. We used the PLUTO 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 represented in the census tract. 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 represented (between 1 and 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])

NUM_LOTS – Total number of unique PLUTO tax lots per CHS ZIP Code.
RES_UNITS – Total number of residential units in all the buildings per CHS ZIP Code.

Breezy Point Tax Lot Value Imputation Process

CHS ZIP Code 11697 lies at the western tip of the Rockaway peninsula, between Rockaway Inlet and Jamaica Bay. The neighborhood is governed by Queens Community Board 14. The community is run by the Breezy Point Cooperative, in which all residents pay the maintenance, security, and community-oriented costs involved with keeping the community private. The cooperative owns the entire 500-acre (2 Km2) community; residents own their homes and hold shares in the cooperative. Within this ZIP Code the majority of the buildings fall within two PLUTO tax lots which both also happen to fall within Breezy Point. However, the PLUTO data variables do not correctly report the number of Residential Units or the Residential and Commercial building areas. These 3 variables are required to calculate the Land Use Mix A and Walkability Index variables. Therefore, their values were imputed.

The process of imputation involved using a GIS layer of building footprints to determine the number of buildings per tax lot. A second GIS layer of building elevations was spatially associated with the building footprints so the elevation of each building could be determined. Next, using Google Street View several buildings were identified simultaneously in Google and in a GIS environment in order to determine an elevation cut-point between 1 and 2 story buildings. Once enough buildings had been identified and objectively compared, a cut-point of 29 feet was used to distinguish 1 and 2 story buildings. The tallest building was 40 feet which was also identified in Google Street View to verify it wasn't a 3 story building, which it was not, suggesting that only 1 and 2 story buildings exist. Then, the square footage of each building was calculated using the building footprints layer and multiplied by the number of stories to determine the total square footage of each building.

To determine commercial vs. residential buildings, Google Street View was again used in combination with prior *in situ* knowledge, and the NumBldgs, UnitsRes, and UnitsTotal PLUTO variables for one of the two tax lots (PLUTO_ID = 311514) as the values appeared to be correct. The values reported that there were 19 commercial buildings. And since each tax lot shares an equal length of road frontage along Rockaway Point Blvd, the road where all of the commercial businesses are believed to be located (based on *in situ* investigation) a value of 19 was attributed to both tax lots as the number of commercial buildings. Finally, building footprints were spatially associated with the two tax lots and the following variables aggregated to each:

- total number of buildings;
- total number of commercial buildings;
- total number of residential buildings;
- total square footage commercial buildings; and
- total square footage residential buildings.

The actual values assigned to each tax lot are as follows:

Tax Lot 1 (PLUTO_ID = 298197)

-NumBldgs: 1816	
-ResUnits: 1797	(1816 - 19)
-BldgArea: 3002833.474657 ft2	(AvgBldgsArea: 1653.542662)
-ResArea: 2971416.164079 ft2	(3002833.474657 - (1653.542662 * 19))
-ComArea: 31417.310578 ft2	(1653.542662 * 19)
-RetailArea: 31417.310578 ft2	(1653.542662 * 19)

Tax Lot 2 (PLUTO_ID = 311514)

-NumBldgs: 597	
-ResUnits: 578	(597 - 19)
-BldgArea: 1131305.939767 ft2	(AvgBldgsArea: 1894.984824)
-ResArea: 1095301.228111 ft2	(1131305.939767 - (1894.984824 * 19))
-ComArea: 36004.711656 ft2	(1894.984824 * 19)
-RetailArea: 36004.711656 ft2	(1894.984824 * 19)

Actual name of the MapPLUTO variables whose values were imputed:
-BldgArea
-ComArea
-ResArea
-RetailArea (same value as ComArea as it is assumed that retail area is the only type of commercial area present)
-NumBldgs
-UnitsRes
-UnitsTotal

ALR Digital Pedestrian Count: Number of People on the Street Estimate

The digital version of the pedestrian count variable was constructed using population data from the U.S. Census for block groups, commercial zoning data from the NYC Department of City Planning's PLUTO database, and subway stop and ridership data from the New York State Department of Transportation and the NYC Transit Authority. These variables were included in this measure because the number of people walking on the street is a factor of the density of the population living in the area, the density of the commercial attractions, and the population passing through on public transportation.

Population density was calculated for all of NYC from block group centroids (geometric center points) using kernel density—with a 100 foot cell size and a search radius of one mile, or 5,280 feet. The kernel density function transforms population values from a series of points to a continuous raster surface with grid cells containing the values for population per square mile. Similarly, commercial zoning density was calculated for all of NYC from tax lot centroids, also using kernel density, a 100 foot cell size and a one mile search radius. The commercially zoned tax lots were first isolated to create a raster grid surface that represented the density of tax lots zoned for commercial uses. Last, subway ridership density was calculated for NYC from subway stop points and ridership values that are associated with these points. Again, the kernel density transformation used a 100 foot cell size and a one mile search radius.

The cell values for each of these three raster grids were then averaged (separately) by the areas of the sampled block polygons. The average values were then transformed to be on scale similar to the pedestrian counts that observers recorded on the street. Population density average values were divided by 10,000, commercial zoning density average values were divided by 100 and subway ridership density average values were divided by 10,000. Last, the three transformed average values were summed for the final digital pedestrian count estimate [ZIP_NUMPEP].

ZIP_POPTOT – Transformed population density average per Mi2 [POPTOT_AVG / 10,000].
ZIP_COMLOT – Transformed commercial zoning density average per Mi2 [COMLOT_AVG / 100].
ZIP_SUBPOP – Transformed population density average per Mi2 [SUBPOP_AVG / 10,000].
ZIP_NUMPEP – Composite ALR digital pedestrian count estimate per Mi2 [ZIP_POPTOT + ZIP_COMLOT + ZIP_SUBPOP].

NYC Park Inspection Program

The Park Inspection Program (PIP) is a comprehensive, outcome-based performance measurement system that generates frequent, random, and detailed inspections of the conditions and quality of the City's public parks, playgrounds, and recreational facilities. Trained inspectors use hand-held computers and digital cameras to conduct nearly 5,000 inspections of ~3,000 properties per year and visit each park once or twice per year. Parks larger than 6 acres are divided into zones and rated separately; playgrounds are also rated separately. The two Figures below uses Central Park of an example of how large parks are divided into zones for PIP rating purposes.

The Table below lists the sixteen features which are split into three categories and which are rated as "Acceptable" or "Unacceptable" during field inspections. The four cleanliness features (litter, glass, weeds, and graffiti) are rated at every park and playground and are used to determine each park's Cleanliness rating of each park or playground. The other twelve features are only rated when available. For example, when rating a park with no amenities only the four Cleanliness features can be rated. However, if a playground is being rated seven or more features are rated, the primary which are: litter, glass, weeds, graffiti, paved surfaces, play equipment, and safety surfaces. All sixteen features are used to determine the Overall Condition of a park. The **CLEANLINESS** condition is rated unacceptable if any two cleanliness features are unacceptable or any one cleanliness feature is very unacceptable. **OVERALL** condition is rated unacceptable if the site fails cleanliness, if any three of any of its features are unacceptable, or if any one of its features is very unacceptable.

To learn more about PIP, please visit: http://www.nycgovparks.org/sub_about/parks_numbers/pip.html.

Figure. – Central Park in Manhattan and its PIP rated zones.

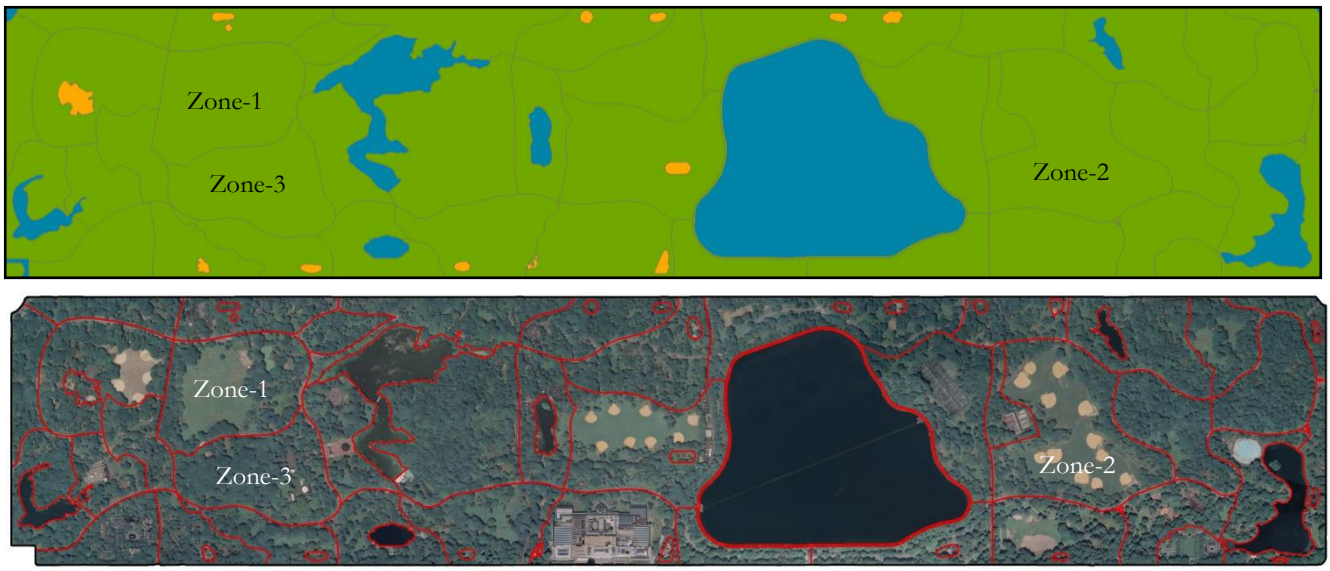


Table. – Sixteen PIP rated park and playground features.

Cleanliness

- Litter
- Glass
- Weeds
- Graffiti

Landscape

- Lawns
- Trees
- Athletic Fields
- Horticultural Areas
- Water Bodies
- Trails

Structural

- Sidewalks
- Paved Surfaces
- Benches
- Fences
- Play Equipment
- Safety Surfaces

Park Inspection Program [PIP] Variables

♠ Variables consider **ALL** park features regardless of the Category or Subcategory Type, **NOTHING** was removed, not Greenstreets not traffic triangle parks, not sitting area park features; nothing.
ALLPARKCNT – Count of **ALL** park features per ZIP Code.
ALLPARKPCT – Percent of ZIP Code occupied by **ALL** park features [ALLPARKKM2 / ZIP_LNDKM2].
ALLPARKKM2 – Land area in Km2 of **ALL** park features per ZIP Code.

♠ Variables consider a reduced number **SELECT** park features, selected based on their Category or Subcategory Type. Park features removed include: "CATMASTER" = ' ' OR "CATMASTER" = 'Greenstreet' OR "CATMASTER" = 'Other' OR "CATMASTER" = 'Water' OR "SCATMASTER" = ' ' OR "SCATMASTER" = 'Cemetery' OR "SCATMASTER" = 'DOT Adopt-A-Highway' OR "SCATMASTER" = 'Golf Course' OR "SCATMASTER" = 'Greenstreet' OR "SCATMASTER" = 'Greenthumb Garden' OR "SCATMASTER" = 'Greenthumb' OR "SCATMASTER" = 'Highway Property' OR "SCATMASTER" = 'Island' OR "SCATMASTER" = 'Natural Area' OR "SCATMASTER" = 'Park Strip' OR "SCATMASTER" = 'Parking Lot' OR "SCATMASTER" = 'Pier' OR "SCATMASTER" = 'Private Property' OR "SCATMASTER" = 'Sitting Area/Triangle/Mall' OR "SCATMASTER" = 'To Be Determined' OR "SCATMASTER" = 'Type 1' OR "SCATMASTER" = 'Type 2' OR "SCATMASTER" = 'Type 3' OR "SCATMASTER" = 'Undeveloped Parkland'. **THIS INCLUDES ALL PARK FEATURES, THAT AVOIDED EXCLUSION FROM THE ABOVE QUERY, WHETHER THEY WERE PIP RATED OR NOT**. [Query Expression: REMOVE_JQ = 1].
SELPARKCNT – Count of **SELECT** park features per ZIP Code.
SELPARKPCT – Percent of ZIP Code occupied by **SELECT** park features [SELPARKKM2 / ZIP_LNDKM2].
SELPARKKM2 – Land area in Km2 of **SELECT** park features per ZIP Code.

♠ Variables consider **ONLY** the PIP rated playgrounds features.
PLGDPIPCNT – Count of **ONLY** the PIP rated playgrounds per ZIP Code.
PLGDPIPPCT – Percent of ZIP Code occupied by **ONLY** PIP rated playgrounds [PLGDPIPKM2 / ZIP_LNDKM2].
PLGDPIPKM2 – Land area in Km2 of **ONLY** the PIP rated playgrounds per ZIP Code.

Count, Area, and Percentage Variables

♠ Variables consider only those park features whose Category Type is Greenstreet, Small Park, Large Park, or Other, regardless of whether they were PIP rated or not PIP rated.
ALL_PRKCNT – Count of all parks [ALL_GRNCNT + ALL_SMLCNT + ALL_LRGCNT + ALL_OTRCNT].
ALL_PRKPCT – Percent of ZIP Code that is covered by all parks.
ALL_PRKKM2 – Land area in Km2 of ZIP Code that is covered by all parks.
ALL_GRNCNT – Count of all greenstreets.
ALL_GRNPCT – Percent of ZIP Code that is covered by all greenstreets.
ALL_GRNKM2 – Land area in Km2 of ZIP Code that is covered by all greenstreets.
ALL_SMLCNT – Count of all small parks [< 6 acres].
ALL_SMLPCT – Percent of ZIP Code that is covered by all small parks.
ALL_SMLKM2 – Land area in Km2 of ZIP Code that is covered by all small parks.
ALL_LRGCNT – Count of all large parks [>= 6 acres].
ALL_LRGPCT – Percent of ZIP Code that is covered by all large parks.
ALL_LRGKM2 – Land area in Km2 of ZIP Code that is covered by all large parks.
ALL_OTRCNT – Count of all "other" park features [not categorized as greenstreets, small or large parks].
ALL_OTRPCT – Percent of ZIP Code that is covered by all "other" park features.
ALL_OTRKM2 – Land area in Km2 of ZIP Code that is covered by all "other" park features.

♠ Variables consider only those park features whose Category Type is Greenstreet, Small Park, Large Park, or Other, AND were PIP rated one or more times between 2002-2006.
PIP_PRKCNT – Count of PIP rated parks [PIP_GRNCNT + PIP_SMLCNT + PIP_LRGCNT + PIP_OTRCNT].
PIP_PRKPCT – Percent of ZIP Code that is covered by PIP rated parks.
PIP_PRKKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated parks.
PIP_GRNCNT – Count of PIP rated greenstreets.
PIP_GRNPCT – Percent of ZIP Code that is covered by PIP rated greenstreets.
PIP_GRNKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated greenstreets.
PIP_SMLCNT – Count of PIP rated small parks [< 6 acres].

PIP_SMLPCT – Percent of ZIP Code that is covered by PIP rated small parks.
PIP_SMLKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated small parks.
PIP_LRGCNT – Count of PIP rated large parks [>= 6 acres].
PIP_LRGPCT – Percent of ZIP Code that is covered by PIP rated large parks.
PIP_LRGKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated large parks.
PIP_OTRCNT – Count of PIP rated "other" park features [not categorized as greenstreets, small or large parks].
PIP_OTRPCT – Percent of ZIP Code that is covered by PIP rated "other" park features.
PIP_OTRKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated "other" park features.

♠ Variable considers only those **park** and **playground** features whose Category Type is Small Park or Large Park AND if they were PIP rated one or more times between 2002-2006.
PARK_PIP – Count of PIP rated park and playground features per ZIP Code used to calculate the PIP rated variables. This count may differ from values found in PIP_PRKCNT because only those park features whose Category Type is Small Park or Large Park are used for PIP variable related analysis and the PIP_PRKCNT variable includes Greenstreets, Small Parks, Large Parks, and Other categories.

♠ The below count values (i.e., PARKPIP_02 - PARKPIP_06) represent, out of the Count available in PARK_PIP, how many times were those total available park and playground features were PIP rated in each year.
PARKPIP_02 – Number of times the available park and playground features were PIP rated per ZIP Code in 2002.
PARKPIP_03 – Number of times the available park and playground features were PIP rated per ZIP Code in 2003.
PARKPIP_04 – Number of times the available park and playground features were PIP rated per ZIP Code in 2004.
PARKPIP_05 – Number of times the available park and playground features were PIP rated per ZIP Code in 2005.
PARKPIP_06 – Number of times the available park and playground features were PIP rated per ZIP Code in 2006.

♠ Variable considers only those park features whose Subcategory Type is DOE Playgrounds, Jointly Operated Playgrounds, Neighborhood Playgrounds, and Playgrounds within Parks…meaning, all **playground** features only ("SCATMASTER" = 'DOE Plgd' OR "SCATMASTER" = 'Jointly Operated Playground' OR "SCATMASTER" = 'Neighborhood Playground' OR "SCATMASTER" = 'Playground Within Park') AND if they were PIP rated one or more times between 2002-2006.
PLGD_PIP – Count of PIP rated playground features per ZIP Code used to calculate the PIP rated variables. This count is the same as those values found in PIP_PRKCNT but is listed again here for the purpose of contiguity.

♠ The below count values (i.e., PLGDPIP_02 - PLGDPIP_06) represent, out of the total count available in PLGD_PIP, how many times were those total available playground features were PIP rated in each year.
PLGDPIP_02 – Number of times the available playground features were PIP rated per ZIP Code in 2002.
PLGDPIP_03 – Number of times the available playground features were PIP rated per ZIP Code in 2003.
PLGDPIP_04 – Number of times the available playground features were PIP rated per ZIP Code in 2004.
PLGDPIP_05 – Number of times the available playground features were PIP rated per ZIP Code in 2005.
PLGDPIP_06 – Number of times the available playground features were PIP rated per ZIP Code in 2006.

Average Park and Playground Rating Score Variables: Individual Years

Using the raw PIPs data from 2002-2006, the seven available park and playground feature inspection ratings (weeds, glass, litter, graffiti, play equipment, safety surfaces and paved areas) were first coded 0 for satisfactory and 1 for unsatisfactory. Next, because park and playground features, can, and usually are, inspected two or more times per year, an average score for each playground feature was calculated across inspections done in a given year (i.e., scale range is still 0-1). Then, CHS merged ZIP Codes were intersected with the park and playgrounds (PIP rated only) in order to identify which park and playground features fell partially or completely within each ZIP Code. The seven available playground ratings were then summed by each unique ZIP and by each year. Finally, the seven PIP ratings that had just been summed by ZIP and year were again averaged for each ZIP and each year, where the numerator was the summed value of each rating and the dominator was the number of times the park and playground features had been rated. The results are variables representing the average PIP quality score for each of the seven measures available for park and playground features.

HIGHER SCORES INDICATE THAT THE PROPERTY DID WORSE; FOR THE SEVEN **INDIVIDUAL** MEASURES, THE AVERAGE SCORE HAVE A POSSIBLE RANGE OF 0-1; THE COMBINED **CLEANLINESS** MEASURE HAS A POSSIBLE RANGE OF 0-4; AND THE COMBINED **QUALITY** MEASURE HAS A POSSIBLE RANGE OF 0-6.

The below variables were calculated using all available PIP rated park and playground features per ZIP Code and are available for 2002-2006, where "YY" in the variable name denotes the year.

WEEDSYYAVG – Average weeds score.
LITERYYAVG – Average litter score.
GLASSYYAVG – Average glass score.
GRAFIYYAVG – Average graffiti score.
PLYEQYYAVG – Average play equipment score.
SAFTYYYAVG – Average safety surfaces score.
PAVDSYYAVG – Average paved surfaces score.
CLNSRYYAVG – Average combined cleanliness score.
SUMSRYYAVG – Average combined quality score.

Average Park and Playground Rating Score Variables: Averaged Across Years

After the seven available park (weeds, glass, litter, and graffiti) and playground feature inspection ratings (weeds, glass, litter, graffiti, play equipment, safety surfaces and paved areas) had been averaged individually by ZIP and year, an average for each of the seven ratings was calculated and averaged across all years. This was accomplished by simply summing the seven ratings for 2002-2006 by ZIP, and then dividing by 5.

HIGHER SCORES INDICATE THAT THE PROPERTY DID WORSE; FOR THE SEVEN **INDIVIDUAL** MEASURES, THE AVERAGE SCORE HAVE A POSSIBLE RANGE OF 0-1; THE COMBINED **CLEANLINESS** MEASURE HAS A POSSIBLE RANGE OF 0-4; AND THE COMBINED **QUALITY** MEASURE HAS A POSSIBLE RANGE OF 0-6.

The below variables were calculated using all available PIP rated park and playground features per ZIP Code.

WEEDS_AVG – Average weeds score.
LITTER_AVG – Average litter score.
GLASS_AVG – Average glass score.
GRAFIT_AVG – Average graffiti score.
PLAYEQ_AVG – Average play equipment score.
SAFETY_AVG – Average safety surfaces score.
PAVEDS_AVG – Average paved surfaces score.
CLNSCR_AVG – Average combined cleanliness score.
SUMSCR_AVG – Average combined quality score.

Minimum–Maximum Park and Playground Rating Score Variables: Individual Years

The minimum and maximum PIP quality scores were calculated for each of the four PIP measures available for parks (weeds, glass, litter, and graffiti) and the seven PIP measures available for playground features (weeds, glass, litter, graffiti, play equipment, paved surfaces, and safety surfaces).

HIGHER SCORES INDICATE THAT THE PROPERTY DID WORSE; FOR THE SEVEN **INDIVIDUAL** MEASURES, THE MINIMUM AND MAXIMUM SCORES HAVE A POSSIBLE RANGE OF 0-1; THE COMBINED **CLEANLINESS** MEASURE HAS A POSSIBLE RANGE OF 0-4; AND THE COMBINED **QUALITY** MEASURE HAS A POSSIBLE RANGE OF 0-6.

The below variables were calculated using all available PIP rated park and playground features per ZIP Code and are available for 2002-2006, where "YY" in the variable name denotes the year.

MINWEEDSYY – Minimum weeds score.
MAXWEEDSYY – Maximum weeds score.
MINLITERYY – Minimum litter score.
MAX LITERYY – Maximum litter score.
MINGLASSYY – Minimum glass score.
MAXGLASSYY – Maximum glass score.
MINGRAFIYY – Minimum graffiti score.
MAX GRAFIYY – Maximum graffiti score.
MINPLYEQYY – Minimum play equipment score.
MAXPLYEQYY – Maximum play equipment score.
MINSAFTYYY – Minimum safety surfaces score.
MAXSAFTYYY – Maximum safety surfaces score.
MINPAVDSYY – Minimum paved surfaces score.
MAXPAVDSYY – Maximum paved surfaces score.
MINCLNSRYY – Minimum combined cleanliness score.
MAXCLNSRYY – Maximum combined cleanliness score.
MINSUMSRYY – Minimum combined quality score.
MAXSUMSRYY – Maximum combined quality score.

Minimum–Maximum Park and Playground Rating Score Variables: Averaged Across Years

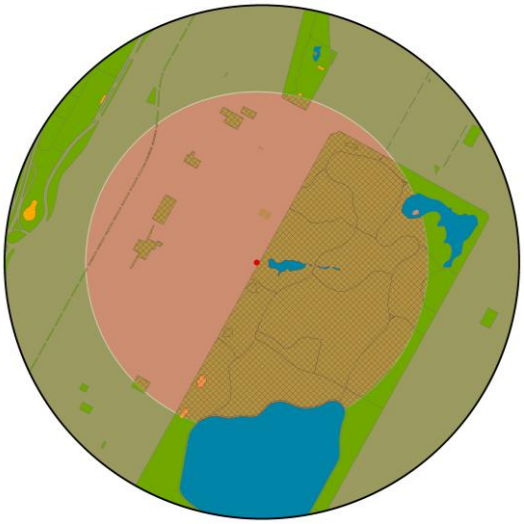
After the minimum and maximum PIP quality scores were calculated for each of the four PIP measures available for parks (weeds, glass, litter, and graffiti) and the seven PIP measures available for playground features (weeds, glass, litter, graffiti, play equipment, paved surfaces, and safety surfaces) individually by ZIP and year, the average for each of the seven ratings was calculated and averaged across all years. This was accomplished by simply summing the seven minimum and maximum ratings for 2002-2006 by ZIP, and then dividing by 5.

HIGHER SCORES INDICATE THAT THE PROPERTY DID WORSE; FOR THE SEVEN **INDIVIDUAL** MEASURES, THE MINIMUM AND MAXIMUM SCORES HAVE A POSSIBLE RANGE OF 0-1; THE COMBINED **CLEANLINESS** MEASURE HAS A POSSIBLE RANGE OF 0-4; AND THE COMBINED **QUALITY** MEASURE HAS A POSSIBLE RANGE OF 0-6.

The below variables were calculated using all available PIP rated park and playground features per ZIP Code.

MIN_WEEDS – Averaged minimum weeds score.
MAX_WEEDS – Averaged maximum weeds score.
MIN_LITTER – Averaged minimum litter score.
MAX_LITTER – Averaged maximum litter score.
MIN_GLASS – Averaged minimum glass score.
MAX_GLASS – Averaged maximum glass score.
MIN_GRAFFI – Averaged minimum graffiti score.
MAX_GRAFFI – Averaged maximum graffiti score.
MIN_PLAYEQ – Averaged minimum play equipment score.
MAX_PLAYEQ – Averaged maximum play equipment score.
MIN_SAFETY – Averaged minimum safety surfaces score.
MAX_SAFETY – Averaged maximum safety surfaces score.
MIN_PAVEDS – Averaged minimum paved surfaces score.
MAX_PAVEDS – Averaged maximum paved surfaces score.
MIN_CSCOR – Averaged minimum combined cleanliness score.
MAX_CSCOR – Averaged maximum combined cleanliness score.
MIN_SSCOR – Averaged minimum combined quality score.
MAX_SSCOR – Averaged maximum combined quality score.

Area-Weighted PIP Cleanliness Score Variables



Using PIPs data from 2002-2006, park inspection features (weeds, glass, litter, graffiti, play equipment, paved surfaces, and safety surfaces) were coded 0 for satisfactory and 1 for unsatisfactory. Within each property, the average score for each feature was calculated across inspections done in 2002-2006 and then averaged across all years. The scale has a range of 0-1 with higher scores being worse. For each property a total cleanliness score and a total quality score was calculated for each year as the sum of the average scores for weeds, glass, litter, graffiti, and again averaged. Higher scores indicate that the property did worse; individual cleanliness scores have a scale ranging from 0-1; the total cleanliness score has a scale ranging from 0-4; and the combined quality score has a scale ranging from 0-6.

Area-Weighted Method – Area-weighted PIP cleanliness score that considers the total area of the park and playground features that fell within each ZIP Code. This PIP area-weighted score considers only the portion of the park you can reach within each ZIP Code and does not factor in the size of the rest of the park outside if it were to across the geographic bounds of two or more ZIP Codes. Therefore, only the portion of a park that is actually within the ZIP Code is considered.

$$WGHT2\_PIP = \frac{\sum_{i=1}^{N} A_i \times R_i \times PIP_i}{\sum_{i=1}^{N} A_i \times R_i}$$

Where: the N parks in the buffer are indexed 1,2,…,N
Ai= total park area of park i
Ri = percentage of park i in buffer
PIPi= PIP cleanliness score of park i

WT2_WEEDS – Method 2: Area-weighted PIP rated weeds score.
WT2_LITTER – Method 2: Area-weighted PIP rated litter score.
WT2_GLASS – Method 2: Area-weighted PIP rated glass score.
WT2_GRAFFI – Method 2: Area-weighted PIP rated graffiti score.
WT2_PLAYEQ – Method 2: Area-weighted PIP rated play equipment score.
WT2_SAFETY – Method 2: Area-weighted PIP rated safety surfaces score.
WT2_PAVEDS – Method 2: Area-weighted PIP rated paved surfaces score.
WT2_CSCOR – Method 2: Area-weighted PIP rated combined cleanliness score.
WT2_SSCOR – Method 2: Area-weighted PIP rated combined quality score.

Park Athletic and Recreational Amenities

Facility counts are per ZIP Code. Athletic and recreational amenities are represented by either point or polygon features. For a point or polygon feature to be counted, it has to simply interest or touch any part of the ZIP Code.

The geographic features used to create these metrics (with exception of access points and playgrounds) were acquired from the NYC Data Mine web-site (www.nyc.gov/data/). Access point and playground data came from the BEH spatial data repository and are modified versions of datasets acquired from Parks and Recreation.

ACCESSPNTS – [points] – Count of park and playground access points.
BASEBALL – [polygons] – Count of baseball fields.
BASKETBALL – [polygons] – Count of basketball courts.
BATHROOMS – [polygons] – Count of open bathrooms.
BEACHZONES – [polygons] – Count of unique beach zones. Beach zones are administrative units designated by the Department of Parks and Recreation. Beaches are split into manageable sizes for the purpose of PIP rating. There are 181 unique beach zones in the City and 7 unique beaches.
BEACHES – [polygons] – Count of beaches.
CALLBOXES – [points] – Count of emergency call boxes.
FLAGPOLES – [points] – Count of flag poles.
GOLFCOURSE – [polygons] – Count of golf courses.
GRNTMBGRDN – [polygons] – Count of green-thumb gardens.
HANDBALL – [polygons] – Count of handball courts.
HOCKEY – [polygons] – Count of hockey rinks.
HOOPS – [points] – Count of basketball hoops (actual point locations of backboard and rim).
MONUMENTS – [points] – Count of monuments.
MULTIFIELD – [polygons] – Count of multi-purpose courts and fields (i.e., several different sports can be played).
PARKINGLOT – [polygons] – Count of park related parking lots.
PICNICAREA – [points] – Count of picnic areas.
PLAYGROUND – [polygons] – Count of playgrounds (includes DOE playgrounds (n=35), jointly operated playgrounds (n=265), neighborhood playgrounds (n=466), and playgrounds within parks (n=269)).
POOLS – [polygons] – Count of open pools.
SOCERFBALL – [polygons] – Count of soccer or "American" football fields.
SHOWERS – [points] – Count of active spray showers.
TENNIS – [polygons] – Count of tennis courts.
TRACKS – [polygons] – Count of running tracks.
VOLLEYBALL – [polygons] – Count of volleyball courts.
WATERFOUNT – [points] – Count of active water fountains.

References

Gattrell, A.C. et al. (1996). Spatial point pattern analysis and its application in geographical epidemiology, *Transactions of the Institute of British Geographers* 21: pp. 256–274.

Guagliardo, M.F. et al. (2004). Physician accessibility: an urban case study of pediatric providers, *Health & Place* 10: pp. 273–283.

Longley, P.A. et al. (2001) *Geographic Information Systems and Science.* Chichester, NY: John Wiley and Sons, Inc.

O'Sullivan, D. and D. Unwin. (2002). *Geographic information analysis.* Hoboken, NJ: John Wiley and Sons, Inc.

Scott, D.W. (1998). *Multivariate Density Estimation*. John Wiley and Sons, Inc.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

Wand, M.P. and M.C. Jones. (1995). *Kernel Smoothing.* London, UK: Chapman and Hall.

DOH Zip Code Merge Decisions

10001 = '10001, 10018, 10120'
10002 = '10002'
10003 = '10003'
10007 = '10007, 10004, 10005, 10006, 10041, 10045, 10278'
10009 = '10009'
10010 = '10010'
10011 = '10011'
10012 = '10012'
10013 = '10013'
10014 = '10014'
10016 = '10016'
10017 = '10017'
10019 = '10019, 10020'
10021 = '10021, 10044, 10162'
10022 = '10022'
10023 = '10023, 10069'
10024 = '10024'
10025 = '10025'
10026 = '10026'
10027 = '10027'
10028 = '10028'
10029 = '10029'
10030 = '10030'
10031 = '10031'
10032 = '10032'
10033 = '10033'
10034 = '10034'
10035 = '10035'
10036 = '10036'
10037 = '10037'
10038 = '10038'
10039 = '10039'
10040 = '10040'
10128 = '10128'
10280 = '10280, 10282'
10301 = '10301'
10302 = '10302'
10303 = '10303'
10304 = '10304'
10305 = '10305'
10306 = '10306'
10307 = '10307'
10308 = '10308'
10309 = '10309'
10310 = '10310'
10312 = '10312'
10314 = '10314'
10451 = '10451'
10452 = '10452'
10453 = '10453'
10454 = '10454'
10455 = '10455'
10456 = '10456'
10457 = '10457'
10458 = '10458'
10459 = '10459, 10474'
10471 = '10471'
10472 = '10472'
10473 = '10473'
10475 = '10475'
11004 = '11004, 11005, 11040, 11001'
11101 = '11101, 11109, 10044'
11102 = '11102'
11103 = '11103'
11104 = '11104'
11105 = '11105'
11106 = '11106'
11201 = '11201'
11203 = '11203'
11204 = '11204'
11205 = '11205'
11206 = '11206'
11207 = '11207'
11208 = '11208'
11209 = '11209, 11252'
11210 = '11210'
11211 = '11211'
11212 = '11212'
11213 = '11213'
11214 = '11214'
11215 = '11215'
11216 = '11216'
11217 = '11217, 11243'
11218 = '11218'
11219 = '11219'
11220 = '11220'
11221 = '11221'
11222 = '11222'
11223 = '11223'
11224 = '11224'
11225 = '11225'
11226 = '11226'
11228 = '11228'
11229 = '11229'
11230 = '11230'
11231 = '11231'
11232 = '11232'
11233 = '11233'
11234 = '11234'
11235 = '11235'
11236 = '11236'
11237 = '11237'
11238 = '11238'
11239 = '11239'
11354 = '11354'
11355 = '11355'
11356 = '11356'
10460 = '10460'
10461 = '10461'
10462 = '10462'
10463 = '10463'
10464 = '10464'
10465 = '10465'
10466 = '10466'
10467 = '10467'
10468 = '10468'
10469 = '10469'
10470 = '10470'
11357 = '11357, 11351'
11358 = '11358'
11360 = '11360, 11359'
11361 = '11361'
11362 = '11362, 11363'
11364 = '11364'
11365 = '11365'
11366 = '11366'
11367 = '11367'
11368 = '11368'
11369 = '11369'
11372 = '11372, 11370'
11373 = '11373'
11374 = '11374'
11375 = '11375'
11377 = '11377'
11378 = '11378'
11379 = '11379'
11385 = '11385'
11411 = '11411'
11412 = '11412'
11413 = '11413'
11414 = '11414'
11415 = '11415'
11416 = '11416'
11417 = '11417'
11418 = '11418'
11419 = '11419'
11420 = '11420'
11421 = '11421'
11422 = '11422'
11423 = '11423'
11426 = '11426'
11427 = '11427'
11428 = '11428'
11429 = '11429'
11432 = '11432'
11433 = '11433'
11434 = '11434'
11435 = '11435'
11436 = '11436, 11430'
11691 = '11691'
11692 = '11692'
11693 = '11693'
11694 = '11694'
11697 = '11697'

BEH Zip Code Merge Decisions

Postal zip codes not present in ZCTA variables:

- 00083 (Central Park)
- 11096 (small piece of zip under JFK that lies mostly in Nassau County)
- 11109 (in Long Island City-no people there)
- 11251 (Naval Station)
- 11359 (Fort Totten)
- 11451 (York College the City University of New York)

Below is a list of the Zip Code merges that were necessary to create the final layer of Zip Codes representative of the 164 Zip Codes present in the ZIPANALYZE variable of the 2002-2006 CHS data. This final Zip Code layer was used to calculate the aforementioned social and built environment variables. In order to create this layer, a dataset of NYC Zip Codes was downloaded from the NYC Data Mine web-site (www.nyc.gov/data/) that is published by DoITT. This dataset of Zip Codes is very extensive and therefore includes contains many small Zip Codes that represent a single building or group of buildings. These types of Zip Codes lie on top of the larger more general Zip Codes. Because of the nature of these Zips, and because they contain no population, they were simply merged into the Zips which surrounds them.

Merges made affecting ZCTA variables as they ARE present in the ZCTA data:

- Deleted 11430 (JFK) (Removed the 4 CHS respondents from ZIP 11430 as defined by ZIP_HOLD variable)
- Deleted 11371 (La Guardia)*
- Deleted Rikers Island portion of 11370 (Census populations were also subtracted prior variable calculations)
- Merged 11001, 11005, and 11040 into 11004
- Merged 10474 into 10459
- Merged 10282 into 10280 (ZIP 10282 wasn't delineated in our older ZIP layer but was in the new DoITT)
- Merged 10044 and 10162 into 10021
- Merged 10018 into 10001
- Merged 10119 into 10001*+
- Merged 10069 into 10023
- Merged 11363 into 11362
- Merged 11370 into 11372
- Merged 10165 into 10017+
- Merged 10167, 10169, 10170, 10171, 10172, 10173, and 10177 into 10017*^
- Merged 10112, 10152, 10153, and 10154 into 10022*^
- Merged 10020 into 10019*
- Merged 10103 and 10111 into 10019*^
- Merged 10115 into 10027*^
- Merged 10271 into 10005*^
- Merged 10041 into 10004* (Lehman caught ZIP 10041 as one of a single building at 55 Water Street)
- Merged 10278 into 10007* (Lehman caught ZIP 10278 as one of a single building at 26 Federal Plaza)
- Merged 10004, 10005, and 10006 into 10007
- Merged 10048 into 10007+ (ZIP 10048 is assigned to the former World Trade Center)

Merges made NOT affecting ZCTA variables as they ARE NOT present in the ZCTA data:

- Deleted 11251
- Deleted 11096
- Merged 11451 into 11433
- Merged 11359 into 11360
- Merged 11109 into 11101 (ZIP 11109 has no population, however, there 13 CHS respondents)
- Merged 10118, 10120, 10121, 10122, and 10123 into 10001^
- Merged 10047 into 10010^
- Merged 10158 into 10016^
- Merged 10166, 10168, 10174, 10175, 10176, and 10178 into 10017^
- Merged 10055, 10151, and 10155 into 10022^
- Merged 10096, 10110, and 10196 into 10036^
- Merged 10097, 10104, 10105, 10106, and 10107 into 10019^
- Merged 10045 into 10038 (Lehman thinks 10045 falls within 10007 but it appears to be in 10038; 1 CHS)
- Merged 10043, 10081, 10203, 10259, 10260, 10265, 10270, 10271, and 10286 into 10005^
- Merged 10275 into 10004^ (Deleted Governors, Ellis, and Liberty Islands)
- Merged 10279 into 10007^
- Merged 10080 into 10048^
- Merged 10285 into 10280^

*ZIP is present in the ZCTA data but has no population.
+A merge that was not discussed amongst the BEH or Lehman groups. The ZIP is small and has a population.
^A merge that was not discussed amongst the BEH or Lehman groups as they are ZIPs that represent a single building or a group of buildings and have no population.

Zip Code feature geometries removed:

- Deleted eight uninhabited islands from ZIP 10314 in Staten Island
- Deleted one uninhabited island from ZIP 10303 in Staten Island
- Deleted four uninhabited islands from ZIP 10307 in Staten Island
- Deleted one uninhabited island from ZIP 10306 in Staten Island
- Deleted one uninhabited island from ZIP 11214 in Brooklyn
- Deleted one uninhabited island from ZIP 11208 in Brooklyn
- Deleted one uninhabited island from ZIP 11239 in Brooklyn
- Deleted two uninhabited island from ZIP 11234 in Brooklyn
- Deleted twenty-four uninhabited islands from ZIP 11693 in Queens
- Deleted two uninhabited islands from ZIP 11357 in Queens
- Deleted one uninhabited island from ZIP 11362 in Queens
- Deleted one uninhabited island from ZIP 11414 in Queens
- Deleted one uninhabited island from ZIP 11422 in Queens
- Deleted one uninhabited island from ZIP 11691 in Queens
- Deleted one uninhabited island from ZIP 10128 in Manhattan
- Deleted one uninhabited island from ZIP 10035 in Manhattan
- Deleted one uninhabited island from ZIP 10044 in Manhattan (just south of Roosevelt Island)
- Deleted one uninhabited island from ZIP 11211 in Manhattan
- Deleted one uninhabited island from ZIP 10475 in The Bronx
- Deleted one uninhabited island from ZIP 10467 in The Bronx
- Deleted one uninhabited island from ZIP 10462 in The Bronx
- Deleted seventeen uninhabited islands from ZIP 10464 in The Bronx
- Deleted two uninhabited islands from ZIP 10465 in The Bronx

Summarized Zip Code Merge Decisions:

August 26, 2010

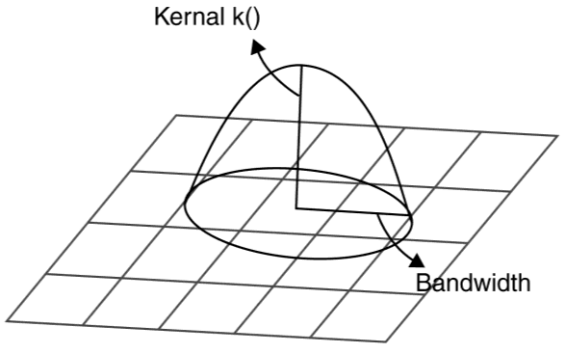
This is a summarized list of ZIP Code merge decisions that does not consider whether ZCTA variables and was assembled for the simple purpose of creating a crosswalk table between NYC ZIP Codes and our CHS ZIP Code merge decisions.

- Deleted 11430 (JFK)
- Deleted 11371 (La Guardia)
- Deleted 00083 (Central Park)
- Deleted 11251 (Naval Station)
- Deleted 11096 (small piece of zip under JFK that lies mostly in Nassau County)

- Merged 11001, 11005, 11040 → 11004
- Merged 10474 →10459
- Merged 10282, 10285 → 10280
- Merged 10044, 10162 → 10021
- Merged 10018, 10118, 10119, 10120, 10121, 10122, 10123 → 10001
- Merged 10069 → 10023
- Merged 11363 → 11362
- Merged 11370 → 11372
- Merged 10165, 10166, 10167, 10168, 10169, 10170, 10171, 10172, 10173, 10174, 10175, 10176, 10177, 10178 → 10017
- Merged 10055, 10112, , 10151, 10152, 10153, 10154, 10155 → 10022
- Merged 10020, 10097, 10103, 10104, 10105, 10106, 10107, 10111 → 10019
- Merged 10115 → 10027
- Merged 10004, 10005, 10006, 10041, 10043, 10048, 10080, 10081, 10203, 10259, 10260, 10265, 10270, 10271, 10275, 10278, 10279, 10286, → 10007
- Merged 11451 → 11433
- Merged 11359 → 11360
- Merged 11109 → 11101
- Merged 10047 → 10010
- Merged 10158 → 10016
- Merged 10096, 10110, 10196 → 10036
- Merged 10045 → 10038

Kernel Density Estimation

The GIS analysis conducted for this project utilized Kernel density estimation (KDE) in order to construct several neighborhood resource facility variables. KDE is a GIS analysis technique that creates a continuous surface based on point data in a neighborhood as defined by a circular distance. Conceptually, a smoothly curved surface is fitted over each point. The surface value is highest at the location of the point and diminishes with increasing distance from the point, reaching zero at the search radius distance from the point (i.e., distance decay function). Density surfaces are effective at identifying where features are concentrated – highlighting areas of intense activity. More a more thorough explanation of KDE, please read the following three paragraphs in this section.



When KDE was employed for this project, a consistent output grid resolution of 150 square feet, a 1-kilometer bandwidth (search radius), and output area units of square kilometers were used. Once the kernel density grids had been created, zonal summary statistics were used to calculate a statistical summary of the values in each raster density grid that fell within the bounds of each Zip Code polygon [min, max, mean, standard deviation, and count]. That is, for each Zip Code, a statistical summary of the values in cells falling within the bounds of each zonal polygon was calculated for inclusion in the models. These zonal aggregations average out any high precision errors that may appear at the individual cell level (Longley et al. 2001). Although several summary statistics are produced by the zonal statistics tool, only average and median values were reported for this analysis.

Kernel Density Description

Kernel density-based approaches create measures that are freed from underlying arbitrarily-bounded administrative units such as Census tracts and zip codes by incorporating distance decay functions. These functions can be used to include information about attributes in a given bandwidth outside a Census tract into the measures generated for each Census tract. This technique uses an arbitrary set of points in a grid overlaid on the study area and uses a three-dimensional floating function to estimate a continuous density surface representing the measure in question (for instance, population density). The density surface is a function of the distances between each grid point and each neighborhood resource facility or Census block group centroid that falls within a predefined distance or bandwidth. The contribution of each neighborhood resource facility to the local density value at a given grid point is based on its distance from the center. Nearby neighborhood resource facilities are given more weight in the density calculation than those farther away. The weights given to neighborhood resource facilities near the boundary defined by the bandwidth approach zero.

In kernel density estimation, the influence of each data point is formally modeled using a mathematical function called a kernel (Silverman 1986, Scott 1998, Wand 1995). When using a kernel function, the population is spread or distributed around the areal unit according to the shape of the kernel. Typical examples of kernels are parabolic, square wave, Gaussian, and the quadratic function. The exact shape or form of the kernel function has a relatively marginal effect on the resulting density estimates (Silverman 1986, Wand 1995). We will use a "Gaussian kernel" method for creating our density grids. This function is the industry standard and is preprogrammed in the ArcGIS Spatial Analyst extension; it utilizes a kernel function based on the quadratic kernel function described in Silverman (1986, p. 76, equation 4.5, Gatrell et al. 1996). This KDE methodology will be used to create continuous surfaces from discrete point data representing two broad categories: neighborhood resource facilities (represented as single points); and Census variables (areal units represented as centroids).

The general form of the kernel density estimator is:

$$w_{i,j} = \begin{cases} \left[1 - \left(\frac{d_{i,j}}{D}\right)^2\right]^2 & d_{i,j} < D \\ 0 \end{cases}$$

where dij is the distance from point i to neighborhood resource facility j and D is the value of the bandwidth. The points, I, represent grid points with a given resolution.

When kernel density estimation is applied, two user-specified parameters can have a dramatic effect on the resulting density: the output resolution of the density grid, and the radius or bandwidth of the kernel function. The grid resolution defines the locations at which estimates are made. (The choice of location of this grid marginally affects the density estimates made, although provided the grid resolution is substantially smaller than the bandwidth the location effect is negligible.) The density estimate is also strongly dependent on the bandwidth parameter of the kernel function: increasing the kernel bandwidth leads to smoothly varying output rasters; decreasing bandwidth produces more localized surface patterns (O'Sullivan and Unwin 2002). It is important to choose a bandwidth that reflects the underlying distribution of the discrete point data (O'Sullivan and Unwin 2002). Research using kernel density functions in dense, urban environments typically uses small grid resolutions and bandwidths to preserve local variation in the underlying Census geographies; decisions are also based on the theoretical understanding of service areas or distances people are likely to travel for a given service (Guagliardo et al. 2004). Using an iterative process, we determined an appropriate grid resolution and bandwidths that maintained the local variation in our data and didn't over-generalize density estimates (O'Sullivan and Unwin 2002).

Appendix A: Path and Country of Birth Codes

POSSIBLE VALUES:
1 = 'Argentina'
2 = 'Australia'
3 = 'Bangladesh'
4 = 'Barbados'
5 = 'Belarus'
6 = 'Bolivia'
7 = 'Brazil'
8 = 'Canada'
9 = 'Caribbean'
10 = 'Chile'
11 = 'China'
12 = 'Colombia'
13 = 'Costa Rica'
14 = 'Cuba'
15 = 'Dominican Republic'
16 = 'Ecuador' (Guayaquil is a city in Ecuador)
17 = 'Egypt'
18 = 'El Salvador'
19 = 'France'
20 = 'Germany'
21 = 'Ghana'
22 = 'Greece'
23 = 'Guatemala'
24 = 'Guyana'
25 = 'Haiti'
26 = 'Honduras'
27 = 'Hong Kong'
28 = 'Hungary'
29 = 'India'
30 = 'Ireland'
31 = 'Israel'
32 = 'Italy'
33 = 'Jamaica'
34 = 'Japan'
35 = 'Korea'
36 = 'Mexico'
37 = 'Nicaragua'
38 = 'Nigeria'

39 = ‘Pakistan’
40 = ‘Panama’
41 = ‘Peru’
42 = ‘Philippines’
43 = ‘Poland’
44 = ‘Puerto Rico’
45 = ‘Romania’
46 = ‘Russia’
47 = ‘Sierra Leone’
48 = ‘South American’
49 = ‘Spain’
50 = ‘Taiwan’
51 = ‘Trinidad and Tobago’
52 = ‘Turkey’
53 = ‘Ukraine’
54 = ‘United Kingdom’
55 = ‘Venezuela’
56 = ‘Vietnam’
57 = ‘West Indian’
58 = ‘Yugoslavia’
59 = ‘Afghanistan’
60 = ‘Africa’
61 = ‘Albania’
62 = ‘Antigua’
63 = ‘Armenia’
64 = ‘Austria’
65 = ‘Bahrain’
67 = ‘Belgium’
68 = ‘Belize’
69 = ‘Bulgaria’
70 = ‘Cambodia’
71 = ‘Central Asia’
72 = ‘Cyprus’
73 = ‘Czechoslovakia’
74 = ‘England’ (see also, code 54)
75 = ‘Europe’
76 = ‘Georgia’ (Also Gruzia)
77 = ‘Unknown’
78 = ‘Grenada’
79 = ‘Guinea’
80 = ‘Indonesia’
81 = ‘Iran’
82 = "Cote D’Ivoire" (‘Ivory Coast’)
83 = ‘Jordan’
84 = ‘Latin American country’
85 = ‘Latvia’
86 = ‘Lebanon’
87 = ‘Malaysia’
88 = ‘Malta’
89 = ‘Middle Eastern’
90 = ‘Montenegro’
91 = ‘Morocco’
92 = ‘Netherlands’
93 = ‘Norway’
94 = ‘Portugal’
95 = ‘Democratic Republic of the Congo (former: Zaire)’
96 = ‘Scotland’ (see also code 54)

97 = 'Senegal'
98 = 'Singapore'
100 = 'South Korea'
101 = 'Sri Lanka'
102 = 'St. Lucia'
103 = 'St. Vincent'
104 = 'Suriname'
105 = 'Sweden'
106 = 'Switzerland'
107 = 'Thailand'
108 = 'Uruguay'
109 = 'Uzbekistan'
110 = 'Virgin Islands'
111 = 'West Africa'
112 = 'Zambia'
113 = 'Dominica' (Official name: The Commonwealth of Dominica-Caribbean Island)
114 = 'Kenya'
115 = 'Alberia' (this is not a country- removed from list 4.10.08)
116 = 'Algeria'
117 = 'Aruba'
118 = 'Asia'
119 = 'Bahamas'
120 = 'Tibet'
121 = 'Bermuda' (British overseas territory)
122 = 'Burma' (Myanmar)
123 = 'British Virgin Islands' (British overseas territory)
124 = 'Cameroon'
125 = 'Central America'
126 = 'Croatia'
127 = 'East Africa'
128 = 'Estonia'
129 = 'Ethiopia'
130 = 'Finland'
131 = 'French Guiana' (French Guiana in S. America)
132 = 'Hawaii' (2 people were coded to Hawaiian in CHS 2003)
133 = 'Iraq'
134 = 'Serbia'
135 = 'Liberia'
136 = 'Lithuania'
137 = 'Mali'
138 = 'Mauritius'
139 = 'Mauritania'
140 = 'Moldova' (Formerly Moldavia)
141 = 'Montserrat ' (Caribbean- British overseas territory)
142 = 'Nepal'
143 = 'New Zealand'
144 = 'Niger'
145 = 'Libya'
146 = 'Martinique' (Tiny island in the Caribbean)
147 = 'Paraguay'
148 = 'Saudi Arabia'
149 = 'South Africa'
150 = 'Southeast Asia'
151 = 'Slovenia'
152 = 'St. Croix'
153 = 'St. Thomas' (part of US Virgin Island)
154 = 'St. Kitts' (should be St Kitts and Nevis-change to code 213 4.10.08)

155 = 'Sudan'
156 = 'Syria'
157 = 'Tajikistan'
158 = 'Uganda'
159 = 'United Arab Emirates'
160 = 'Yemen'
161 = 'Zimbabwe'
162 = 'Angola'
163 = 'Anguilla' (British overseas territory)
164 = 'Tanzania'
165 = 'Azerbaijan'
166 = 'Slovakia'
167 = 'Bosnia'
168 = 'Somalia'
169 = 'Brunei'
170 = 'Cayman Islands'
171 = 'Curacao'
172 = 'Denmark'
173 = 'Gambia'
174 = 'Iceland'
175 = 'Kuwait'
176 = 'Macauo' (Southeast Asia)
177 = 'Nevis' (Caribbean Island-Should be St Kitts and Nevis change to code 213 4.10.08)
178 = 'Palestine'
179 = 'Papua, New Guinea.' (East of Indonesia)
180 = 'Rwanda' (Central Africa)
181 = 'Scandinavia'
182 = 'Montenegro' (duplicate see 90)
183 = 'St. Christopher' former name of St Kitts, change to code 213 4.10.08)
184 = 'St. Martin' (French West Indies)
185 = 'The Island of Dominica' (Commonwealth of Dominica- repeat of code 113, delete this one 4.10.08)
186 = 'Togo'
187 = 'Sicily' (should be coded to Italy- remove from code 4.10.08)
188 = 'St. Maarten' (changed to correct spelling 4.10.08- 'St. Martin')
189 = 'US Virgin Islands'
190 = 'Republic of Macedonia'
191 = 'Kazakhstan'
192 = 'Kyrgyzstan'
193 = 'Gabon' (Central Africa)
194 = 'Tunisia'
195 = 'TurKmenistan'
196 = 'Guam'
197 = 'Eritrea' (East Africa)
198 = 'Fiji'
199 = 'Macedonia' (Repeat of code 190, remove 4.10.08)
200 = 'Samoa'
201 = 'Siberia' (not a country- code to Russia 4.10.08)
202 = 'Turks and Caicos Island' (Caribbean- British overseas territory)
203 = 'Wales (UK)' (see also code 54)
204 = 'Djibouti (East Africa)'
205 = 'St. Troy' (not a place- maybe meant St Croix? remove from code 4.10.08)
206 = 'St. John's' (US VI)
207 = 'Cape Verde' (Group of Islands near West Africa)
208 = 'Colombia' (reassigned to format 12 - Colombia, 3.6.08)
209 = 'Bhutan' (Kingdom of Bhutan got independence from India in 1949)
210 = 'Luxembourg' (It is in Europe)
211 = 'Tadzykistan' (spelled Tajikstan-see 157)

212 = 'Kazbekistan' (THIS IS NOT A COUNTRY)
213 = 'Federation of St Kitts and Nevis'
214 = 'Burkina Faso'
215 = 'Burundi'
216 = 'Kosovo' (Recognized as independent state http://www.state.gov/p/eur/ci/kv/)
217 = 'Guadeloupe'
218 = 'Namibia'
219 = 'Former USSR, not specified'
1000 = 'USA'

^**ATHOMELANG** (Athomelang567p) – What language do you speak most often at home?

POSSIBLE VALUES:
- 1 = 'English'
- 2 = 'Spanish'
- 3 = 'Other'

^**ATHOMELAN0** (Athomelang67p) – What language do you speak most often at home?

POSSIBLE VALUES:
- 1 = 'English'
- 2 = 'Spanish'
- 3 = 'Russian'
- 4 = 'Chinese'
- 5 = 'Indian'
- 6 = 'Other'

GIS Measures of the Built and Social Environment

Below is a list and brief description of the social and built environment variables that were calculated for the CHS Zip Codes. Please use this document as a data dictionary for the accompanying tables.

ZIP – Zip Code variable used for GIS processing and unique table joining feature key ID.
ZIP_KM2 – Total area of Zip Code in Km2.
ZIP_LNDKM2 – Total land area of Zip Code in Km2 (inland water features subtracted out). THIS VARIABLE SHOULD BE USED WHEN CALCULATING DENSITY ESTIMATION VARIABLES.

Social Environment Variables

US Census Summary File 3 Variables

♠ ALL 2000 US CENSUS VARIABLES WERE CALCULATED USING VARIABLES FROM SUMMARY FILE 3 [SF3] SAMPLE DATA <www.census.gov/prod/cen2000/doc/sf3.pdf>. WHEN POSSIBLE, VARIABLES WERE FIRST CALCULATED USING BLOCK GROUP LEVEL VARIABLES. HOWEVER, THERE ARE CENSUS VARIABLES AS PART OF THIS GEO-SPATIAL ANALYSIS, WHICH WERE ONLY AVAILABLE AT THE CENSUS TRACT LEVEL [I.E., LANGUAGE SPOKEN AT HOME VARIABLES].

Age Variables

ZIP_TOTPOP – Total population [P001001].
ZIP_POPDEN – Population density per Km2 [P001001] / [ZIP_LNDKM2].
ZIP_TOTMAL – Total population: Male [P008002].
ZIP_TOTFEM – Total population: Female [P008041].

Population & Race Variables

ZIP_CNTWHT – Count of the White population (from the 'single race' variable) [P006002].
ZIP_PCTWHT – Percent White population (from the 'single race' variable) [P006002] / [P006001].
ZIP_CNTBLK – Count of the Black or African American population (from the 'single race' variable) [P006003].
ZIP_PCTBLK – Percent Black or African American population [P006003] / [P006001].
ZIP_CNTLAT – Count of the Hispanic or Latino population (from Hispanic or Latino by Race variable) [P007010]
ZIP_PCTLAT – Percent of the Hispanic or Latino population [P007010] / [P007001].
ZIP_CNTNAT – Count of the American Indian or Alaska Native (includes: American Indian, American Indian tribe, Alaska Native) population (from the 'single race variable) [P006004].

ZIP_PCTNAT – Percent American Indian or Alaska Native (includes: American Indian, American Indian tribe, Alaska Native) population [P006004] / [P006001].
ZIP_CNTASN – Count of the Asian (includes: Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, Other Asian) population (from the 'single race variable) [P006005].
ZIP_PCTASN – Percent of the Asian (includes: Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, Other Asian) population [P006005] / [P006001].
ZIP_CNTPAC – Count of the Native Hawaiian or Other Pacific Islander (includes: Native Hawaiian, Guamanian, Chamorro, Samoan, Other Pacific Islander—Carolinian ; Chuukese (Trukese); Fijian; Kosraean; Melanesian; Micronesian; Northern Mariana Islander; Palauan; Papua New Guinean; Pohnpeian; Polynesian; Solomon Islander; Tahitian; Tokelauan; Tongan; Yapese) population (from the 'single race' variable) [P006006].
ZIP_PCTPAC – Percent of the Native Hawaiian or Other Pacific Islander (includes: Native Hawaiian, Guamanian, Chamorro, Samoan, Other Pacific Islander—Carolinian ; Chuukese (Trukese); Fijian; Kosraean; Melanesian; Micronesian; Northern Mariana Islander; Palauan; Papua New Guinean; Pohnpeian; Polynesian; Solomon Islander; Tahitian; Tokelauan; Tongan; Yapese) population [P006006] / [P006001].
ZIP_CNTOTR – Count of the Some other race population (from the 'single race' variable) [P006007].
ZIP_PCTOTR – Percent of the Some other race population [P006007] / [P006001].
ZIP_CNTTWO – Count of the Two or more races population (from the 'single race' variable) [P006008].
ZIP_PCTTWO – Percent of the Two or more races population [P006008] / [P006001].
ZIP_FORBRN – Total foreign born population [P021013].
ZIP_FORDEN – Density of foreign born population per Km2 [P021013 / ZIP_LNDKM2].
ZIP_PCTFOR – Percent of foreign born population [P021013] / [P001001]].
ZIP_PCTPOV – Percent of total population for whom poverty status is determined [no institutionalized population] whose income is below 100% of the federal poverty line [((P088002 + P088003 + P088004) / P088001)].
ZIP_MHHINC – Median household income in 1999 [dollars] from US Census sf3 [P053001]. Unlike most other US Census variables, this variable is not a count value but rather a median. Therefore, rather than apportioning the Census Block Groups where they are cut the ZIP Code measurement geographies, the median household income value for every Block Group that intersected each ZIP was considered and the median value of all those considered values was calculated.

Household Variables

ZIP_HHTOT – House-holds: Total [P020001].
ZIP_HHSOC – House-holds: With Social Security income [P062002].
ZIP_HHSSI – House-holds: With Supplemental Security Income (SSI) [P063002].
ZIP_HHPAI – House-holds: With public assistance income [P064002].

Housing Unit Variables

ZIP_HUTOT – Housing units: Total [H006001].
ZIP_HUOCCP – Occupied housing units: Total [H006002].
ZIP_PCTOCP – Percent of housing units that are occupied [H006002 / H006001].
ZIP_VACANT – Vacant housing units: Total [H006003].
ZIP_PCTVCT – Percent of house units that are vacant [H006003 / H006001].
ZIP_HUOWN – Owner occupied housing units: Total [H007002].
ZIP_PCTOWN – Percent of occupied house-holds that are owned [H007002 / H007001].
ZIP_HURENT – Renter occupied housing units: Total [H007003].
ZIP_PCTRNT – Percent of occupied house units that are rented [H007003 / H007001].

Commute to Work Variables

ZIP_PCTCAR – Percent of workers 16 years & over whom commuted to work using a private vehicle. Means of a private vehicle include cars; trucks; vans; carpooling; and driving alone [P030002 / P030001].
ZIP_PCTPUB – Percent of workers 16 years & older who commuted to work using public transportation. Means of public transportation include, bus or trolley bus; streetcar or trolley car; subway or elevated; railroad; ferryboat; and taxicab [P030005 / P030001].
ZIP_PCTWLK – Percent of workers 16 years & older who commuted to work by walking [P0300014/ P030001].
ZIP_PCTBIK – Percent of workers 16 years & older who commuted to work using a bicycle [P030013 / P030001].

Work Related Variables

The 'Work Related Variables' (P26, P27, and P30-P35) were added as a result of a discussion during the BEH Team Meeting on Tuesday, August 10th about how far people travel to get to work and by what means of transportation. These variables were identified as being relevant to that discussion and are all simple count variables representing the total number of workers (i.e., persons) or travel time to work in **minutes**. No percentage (%) variables were calculated for the simple reason that the numbers of variables presented in this section were many and it was unknown which variables would be of interest for analytical purposes. However, percentage variables can easily be calculated using the variable of interest as the numerator and the **total** population value from the respective census category as the denominator. E.G., to calculate the 'percent of workers who worked in their county of residence' you would use the following expression: (P026003 / P026001]. Variables under the P30 category in red are flagged as they were previously used to calculate several variables (ZIP_PCTCAR, ZIP_PCTPUB, ZIP_PCTWLK, and ZIP_PCTBIK).

P26. Place of Work for Workers 16 Years and Over--State and County Level

Universe: Workers 16 years and over

Total:	**P026001**
Worked in state of residence:	P026002
Worked in county of residence	P026003
Worked outside county of residence	P026004
Worked outside state of residence	P026005

P27. Place of Work for Workers 16 Years and Over--Place Level

Universe: Workers 16 years and over

Total:	**P027001**
Living in a place:	P027002
Worked in place of residence	P027003
Worked outside place of residence	P027004
Not living in a place	P027005

P30. Means of Transportation to Work for Workers 16 Years and Over

Universe: Workers 16 years and over

Total:	**P030001**
Car, truck, or van:	P030002
Drove alone	P030003
Carpooled	P030004
Public transportation:	P030005
Bus or trolley bus	P030006
Streetcar or trolley car	P030007
Subway or elevated	P030008
Railroad	P030009
Ferryboat	P030010
Taxicab	P030011
Motorcycle	P030012
Bicycle	P030013
Walked	P030014
Other means	P030015
Worked at home	P030016

P31. Travel Time to Work for Workers 16 Years and Over

Universe: Workers 16 years and over

Total:	**P031001**
Did not work at home:	P031002
Less than 5 minutes	P031003
5 to 9 minutes	P031004
10 to 14 minutes	P031005
15 to 19 minutes	P031006
20 to 24 minutes	P031007
25 to 29 minutes	P031008
30 to 34 minutes	P031009
35 to 39 minutes	P031010
40 to 44 minutes	P031011
45 to 59 minutes	P031012
60 to 89 minutes	P031013
90 or more minutes	P031014
Worked at home	P031015

P32. Travel Time to Work by Means of Transportation to Work for Workers 16 Years and Over Who Did Not Work at Home

Universe: Workers 16 years and over who did not work at home

Total:	**P032001**
Less than 30 minutes:	P032002
Public transportation	P032003
Other means	P032004
30 to 44 minutes:	P032005
Public transportation	P032006
Other means	P032007
45 to 59 minutes:	P032008
Public transportation	P032009
Other means	P032010
60 or more minutes:	P032011
Public transportation	P032012
Other means	P032013

P33. Aggregate Travel Time to Work (in minutes) by Travel Time to Work by Means of Transportation to Work for Workers 16 Years and Over Who Did Not Work at Home

Universe: Workers 16 years and over who did not work at home

Aggregate travel time to work (in minutes)**:**	**P033001**
Less than 30 minutes:	P033002
Public transportation	P033003
Other means	P033004
30 to 44 minutes:	P033005
Public transportation	P033006
Other means	P033007
45 to 59 minutes:	P033008
Public transportation	P033009
Other means	P033010
60 or more minutes:	P033011
Public transportation	P033012
Other means	P033013

P34. Time Leaving Home to Go to Work for Workers 16 Years and Over

Universe: Workers 16 years and over

Total:	**P034001**
Did not work at home:	P034002
12:00 a.m. to 4:59 a.m.	P034003
5:00 a.m. to 5:29 a.m.	P034004
5:30 a.m. to 5:59 a.m.	P034005
6:00 a.m. to 6:29 a.m.	P034006
6:30 a.m. to 6:59 a.m.	P034007
7:00 a.m. to 7:29 a.m.	P034008
7:30 a.m. to 7:59 a.m.	P034009
8:00 a.m. to 8:29 a.m.	P034010
8:30 a.m. to 8:59 a.m.	P034011
9:00 a.m. to 9:59 a.m.	P034012
10:00 a.m. to 10:59 a.m.	P034013
11:00 a.m. to 11:59 a.m.	P034014
12:00 p.m. to 3:59 p.m.	P034015
4:00 p.m. to 11:59 p.m.	P034016
Worked at home	P034017

P35. Private Vehicle Occupancy for Workers 16 Years and Over

Universe: Workers 16 years and over

Total:	**P035001**
Car, truck, or van:	P035002
Drove alone	P035003
Carpooled:	P035004
In 2-person carpool	P035005
In 3-person carpool	P035006
In 4-person carpool	P035007
In 5- or 6-person carpool	P035008
In 7-or-more-person carpool	P035009
Other means (including those who worked at home)	P035010

Linguistic Isolation Variables

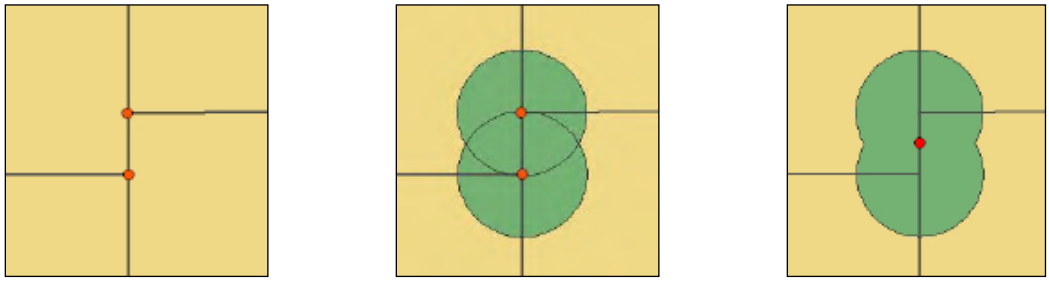
ZIP_HHENGL – Count of *English* speaking house-holds by linguistic isolation [P020002].
ZIP_HHSPAN – Count of *Spanish* speaking house-holds linguistic isolation [P020003].
ZIP_PCTLIN – Percent of population who are linguistically isolated [(P020004 + P020007 + P020010 + P020013) / P020001)].
ZIP_LINSPH – Count of house-holds that are linguistically isolated that speak *Spanish* [P020004].
ZIP_PSPLIN – Percent of house-holds that are linguistically isolated that speak *Spanish* [P020004] / [P020003] (total Spanish house-holds).
ZIP_LINEUR – Count of house-holds that are linguistically isolated that speak *Other Indo-European* languages [P020007].
ZIP_PIELIN – Percent of house-holds that are linguistically isolated that speak *Other Indo-European* languages [P020007] / [P020006] (total Other Indo-European house-holds).
ZIP_LINASN – Count of house-holds that are linguistically isolated that speak *Asian and Pacific Island* languages [P020010].
ZIP_PASLIN – Percent of house-holds that are linguistically isolated that speak *Asian and Pacific Island* languages [P020010] / [P020009] (total Asian and Pacific Island house-holds).
ZIP_LINOTR – Count of house-holds that are linguistically isolated that speak *Other* languages [P020013].
ZIP_POTLIN – Percent of house-holds that are linguistically isolated that speak *Other* languages [P020013] / [P020012] (total Other languages house-holds).

Built Environment Variables

Street Pattern Variables

ZIP_STRINT – Count of "unique" street intersections per ZIP Code 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.). The "unique" count takes into consideration that false positives can be created in regards to total intersections counts since judging an intersection can be complex, particularly in older areas where intersections are not regularly spaced, road rights of way have varying sizes, and centerlines may be slightly offset within the right of way. Therefore, I played with buffering each intersection with several different buffer sizes to help to deal with slightly offset street centerlines and divided roads (e.g., Broadway). Such situations result in multiple intersections where in reality they should be perceived as one intersection. After testing different sized buffered, I concluded that I needed to used 20-meter buffers around divided street nodes and 10-meter buffers around all other street nodes. After the buffers are created, they are merged where they overlap, boundaries are dissolved, and a centroid is created that represents the new street intersection. The figure below shows an example of where such methods would be used. This measure was calculated using the street centerline GIS layer from the New York State Accident Location Information System (NYS-ALIS). Prior to calculating this measure all primary highways with limited access and access ramps were removed since these street features are not used by pedestrians [ZIP_3WYINT + ZIP_4WYINT].

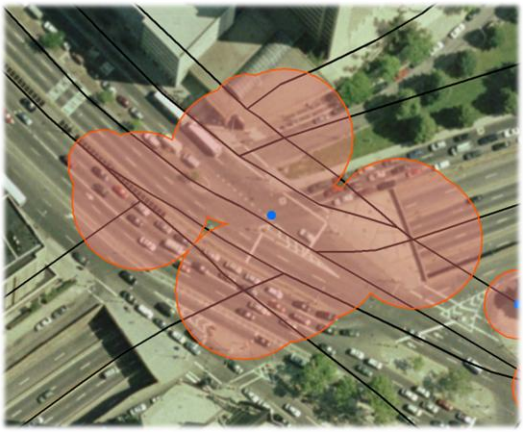
Figure. – Street intersection cleaning process.

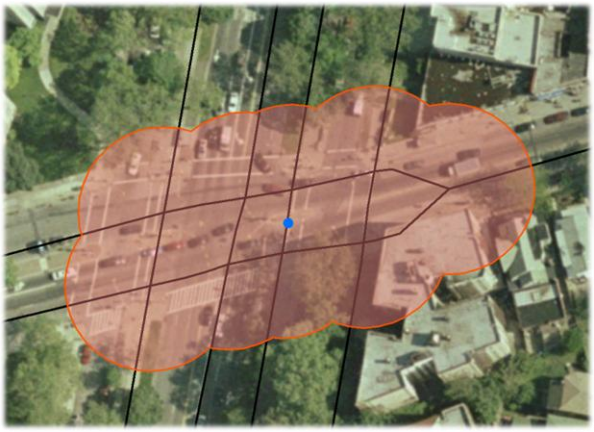


before buffering | after buffering by 10-meters | after merging, dissolving, and creating new centroid

ZIP_INTDEN – Density of "unique" streets intersections per ZIP Code in Km2 [ZIP_STRINT / LNDKM2].
ZIP_CULEND – Count of dead-ends and cul-de-sacs per ZIP Code [valence = 1].
ZIP_CULDEN – Density of dead-ends and cul-de-sacs per ZIP Code in Km2 [ZIP_CULEND / ZIP_LNDKM2].
ZIP_3WYINT – Count of three-way intersections per ZIP Code [valence = 3].
ZIP_3WYDEN – Density of three-way intersections per ZIP Code in Km2 [ZIP_3WYINT / ZIP_LNDKM2].
ZIP_3WYPCT – Percent of three-way intersections to all intersections = raw percentage of three-way intersections versus other intersections [valence 3+], with the added complication that some higher valence intersections are actually three-way and four-way intersections [ZIP_3WYINT / ZIP_STRINT].
ZIP_4WYINT – Count of four-way intersections per ZIP Code [valence = 4]. Four-way intersections are an indicator of grid street patterns, thought to be supportive for walking.
ZIP_4WYDEN – Density of four-way intersections per ZIP Code in Km2 [ZIP_4WYINT / ZIP_LNDKM2].
ZIP_4WYPCT – Percent of four-way intersections to all intersections = raw percentage of four-way intersections versus other intersections [valence 3+], with the added complication that some higher valence intersections are actually three-way and four-way intersections. Four-way intersections are an indicator of grid street patterns, believed by some to be more connected. One can imagine a high proportion of four-way intersections in an area, however, with enormous blocks, so the density of intersections should also be considered [ZIP_4WYINT / ZIP_STRINT].

PLEASE NOTE: As outlined in the discussion of the **ZIP_STRINT variable above, four-way and three-way intersection counts and percentages do not include intersections formed by primary highways with limited access and access ramps since these street features are not used by pedestrians, and counts reflect roads corrected for misaligned intersections by creating buffers around each intersection node, merging them where they overlapped, and creating a centroid for each new intersection polygon. See figures below for such examples.





ZIP_CNR – Connected node ratio is the number of street intersections (valence 3+) divided by the number of intersections plus cul-de-sacs (valence 1) (Dill 2003, 3) per ZIP Code. A connected node ratio close to 1.00 (the maximum value) means that there are few cul-de-sacs and, therefore, the street pattern is more highly connected. Dill (2003, 3) reports that ratios of 0.70 are preferred by those promoting highly-connected street patterns and recommends against networks with values less than 0.5. Of course there could be a perfect grid (CNR = 1.00(with enormous blocks that were hard to get around and such a pattern would fare well on this measure, so it is not a perfect indicator [ZIP_STRINT / [ZIP_CULEND + ZIP_STRINT]].

ZIP_STRKM – Total length of streets per ZIP Code in Km. Total street length was derived using the same shapefile described for the below ZIP_STRWID variable, using only those street centerline feature, where: FEATURE_CO = 2900 (Paved Road).
ZIP_STRDEN – Density of streets per ZIP Code in Km2 [ZIP_STRKM / ZIP_LNDKM2].
ZIP_STRKM2 – Total area of ZIP Code covered by street-network area (i.e., curb-to-curb roadbed) in Km2.
ZIP_STRPCT – Percent of total measurement geography covered by street-network area (i.e., curb-to-curb roadbed). The dark green areas in the figure to the right represents the total area of the street network while the orange are building footprint and the light green/tan is open area. To calculate the percentage street-network area, the street-network area layer was intersected with the different measurement geographies and the ratio of area occupied by street-network area calculated as a ratio [ZIP_STRKM2 / ZIP_LNDKM2].

ZIP_STRWID – Average street width per ZIP Code in Km. To determine the average street width the Streets Centerline layer was acquired from DoITT and all streets, where: FEATURE_CO = 2900 (Paved Road) where exported to a new shapefile. The newly exported shapefile was then intersected with the measurement geographies, roads dissolved by their unique and the average street width was derived using the available WIDTH variable (which is in feet but was converted to kilometers).

- DoITT GIS Web-Site: http://www.nyc.gov/html/doitt/html/eservices/eservices_gis_downloads.shtml
- Possible FEATURE_CO values:
  - 2245: Boardwalk
  - 2265: Interior Path
  - 2900: Paved Road
  - 2910: Unpaved Road
  - 2920: Alley
  - 2930: Stepped Street
  - 2240: Driveway

ZIP_STRMED – Median street width per ZIP Code in Km. The variable was calculated using the same DoITT Streets Centerline layer used to calculate the average street width variable.

Functional Road Classification Variables

Functional classifications of roads divide them into a hierarchy including subclasses of the road types described in detail following the listed variables. These measures assess the total length and ratio of the aforementioned functional classification road classes. Using the FCCs field from the road layers that were intersected with the network buffers, road segments of varying subclasses were summed and divided by the sum of the lengths of all roads to obtain the proportion of different subclass roads to all other roads.

ZIP_A1_KM – Total length in Km of **A1** – Primary Highways with Limited Access roads.
ZIP_A1_PCT – Total percentage of **A1** – Primary Highways with Limited Access roads.
ZIP_A2_KM – Total length in Km of **A2** – Primary Roads without Limited Access.
ZIP_A2_PCT – Total percentage of **A2** – Primary Roads without Limited Access.
ZIP_A3_KM – Total length in Km of **A3** – Secondary and Connecting Roads.
ZIP_A3_PCT – Total percentage of **A3** – Secondary and Connecting Roads.
ZIP_A4_KM – Total length in Km of **A4** – Local, Neighborhood, and Rural Roads.
ZIP_A4_PCT – Total percentage of **A4** – Local, Neighborhood, and Rural Roads.
ZIP_A5_KM – Total length in Km of **A5** – Vehicular Trails.
ZIP_A5_PCT – Total percentage of **A5** – Vehicular Trails.
ZIP_A6_KM – Total length in Km of **A6** – Roads with Special Characteristics.
ZIP_A6_PCT – Total percentage of **A6** – Roads with Special Characteristics.
ZIP_A7_KM – Total length in Km of **A7** – Roads as Other Thoroughfare.
ZIP_A7_PCT – Total percentage of **A7** – Roads as Other Thoroughfare.

- **A1 – Primary Highways With Limited Access—mostly interstates, include:**
  - **A10** Primary road with limited access, major category
  - **A11** Primary road with limited access or interstate hwy, unseparated
  - **A12** Primary road with limited access or interstate hwy, unseparated in tunnel
  - **A13** Primary road with limited access or interstate hwy, unseparated underpassing
  - **A14** Primary road with limited access or interstate hwy, unseparated rail line in center
  - **A15** Primary road with limited access or interstate hwy, separated
  - **A16** Primary road with limited access or interstate hwy, separated in tunnel
  - **A17** Primary road with limited access or interstate hwy, separated underpassing
  - **A18** Primary road with limited access or interstate hwy, separated rail line in center

- **A2 – Primary Roads Without Limited Access—mostly US and state highways, include:**
  - **A20** Primary Highways without limited access, major category
  - **A21** Primary Highways without limited access, unseparated
  - **A22** Primary Highways without limited access, unseparated in tunnel
  - **A23** Primary Highways without limited access, unseparated underpassing
  - **A24** Primary Highways without limited access, unseparated rail line in center
  - **A25** Primary Highways without limited access, separated
  - **A26** Primary Highways without limited access, separated in tunnel
  - **A27** Primary Highways without limited access, separated underpassing
  - **A28** Primary Highways without limited access, separated rail line in center

- **A3 – Secondary and Connecting Roads—mostly state and some county highways, include:**
  - **A30** Secondary state and county highways, major category
  - **A31** Secondary state and county highways, unseparated
  - **A32** Secondary state and county highways, unseparated in tunnel
  - **A33** Secondary state and county highways, unseparated underpassing
  - **A34** Secondary state and county highways, unseparated rail line in center
  - **A35** Secondary state and county highways, separated

A36 Secondary state and county highways, separated in tunnel
A37 Secondary state and county highways, separated underpassing
A38 Secondary state and county highways, separated rail line in center

- **A4 – Local, Neighborhood, and Rural Roads, include:**
  - **A40** Local, neighborhood, rural road, city-street, major category
  - **A41** Local, neighborhood, rural road, city-street, unseparated
  - **A42** Local, neighborhood, rural road, city-street, unseparated in tunnel
  - **A43** Local, neighborhood, rural road, city-street, unseparated underpassing
  - **A44** Local, neighborhood, rural road, city-street, unseparated rail line in center
  - **A45** Local, neighborhood, rural road, city-street, separated
  - **A46** Local, neighborhood, rural road, city-street, separated in tunnel
  - **A47** Local, neighborhood, rural road, city-street, separated underpassing
  - **A48** Local, neighborhood, rural road, city-street, separated rail line in center

- **A5 – Vehicular Trails, include:**
  - **A50** Vehicular (4WD) Trail, major category
  - **A51** Vehicular (4WD) Trail, unseparated
  - **A52** Vehicular (4WD) Trail, unseparated in tunnel
  - **A53** Vehicular (4WD) Trail, unseparated underpassing

- **A6 – Roads with Special Characteristics, include:**
  - **A60** At-grade ramp or connecting road not associated with a limited access highway
  - **A61** Cul-de-Sac, the closed end of a road that forms a loop or turn around
  - **A62** Traffic Circle, the portion of a road or intersection of roads forming a roundabout
  - **A63** Access Ramp, the portion of a road that forms a cloverleaf or limited access interchange
  - **A64** Service Road, provides access to businesses and rest areas
  - **A65** Ferry Crossing, Passenger, Seasonal
  - **A66** Ferry Crossing, Passenger, Year-Round
  - **A67** *not used*
  - **A68** Ferry Crossing, Vehicular, Seasonal
  - **A69** Ferry Crossing, Vehicular, Year-Round

- **A7 – Roads as Other Thoroughfare, include:**
  - **A70** Other Thoroughfare major category
  - **A71** Walkway, for pedestrians, usually unnamed
  - **A72** Stairway, stepped road for pedestrians, usually unnamed
  - **A73** Alley, road for service vehicles, located at the rear of buildings
  - **A74** Driveway, usually privately owned and unnamed
  - **A75** Road, parking area

Speed Limit Variable: Length-Weighted

MPHWGHTAVG – Length-weighted average speed limit per ZIP Code in miles per hour (mph).

$$WGHT\_SPEED = \frac{\sum_{j=1}^{N} LENGTH(S(j)) \times S(j)}{\sum_{j=1}^{N} LENGTH(S(j))}$$

Where: the N possible speed limits in the dataset are indexed 1,2,…,N (in this case N=7)

S = {1, 15, 25, 35, 45, 55, 65} is the set of possible speed limits (e.g. S(3) = 25)

LENGTH(S(j))= length in Km of roads with speed limit S(j) in buffer

Average Daily Traffic Variable: Length-Weighted

ADTWGHTAVG – Length-weighted average daily traffic per ZIP Code in number of vehicles.

Public Transportation Variables

♠ The New York City **Bus** information was originally obtained from the New York City Transit Authority in 1998. The data was realigned to the NYCMap basemap in 2002, and again updated in 2004.

BUS04_CNT – Count of unique MTA bus stops per ZIP Code as of 2004. This measure takes into account those bus stops that have multiple route opportunities at one location and each stop is counted only once regardless of how many route opportunities are available at any given bus stop.
BUS04_DEN – Density of unique bus stops per Km2 [BUS04_CNT / ZIP_LNDKM2].
BUS04_AVG – Kernel Density Estimation variable: **average** number of unique bus stops per Km2.
BUS04_MED – Kernel Density Estimation variable: **median** number of unique bus stops per Km2.

♠ 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.

SUBW07_CNT – Count of MTA subway stations per ZIP Code as of 2007.
SUBW07_DEN – Density of subway stations per Km2 [SUBW07_CNT /ZIP_LNDKM2].
SUBW07_AVG – Kernel Density Estimation variable: **average** number of subway stops per Km2.
SUBW07_MED – Kernel Density Estimation variable: **median** number of subway stops per Km2.

♠ Please note that the GIS Analyst is aware of the fact that the CHS data used for this project represents 2002-2006 and that the spatial dataset of NYC subway stations is from 2007. However, because subway station locations in NYC very, very rarely change, and a decision was made to use the 2007 subway station dataset because of its highly accurate map scale of 1:2,400. The accuracy of this dataset far exceeds the spatial precision of any other subway stations dataset in the BEH spatial data library.

Bicycle Route Variables: 2002 & 2007

New York City's bicycle network is a series of existing, proposed, and planned bicycle routes. These three categories can further be broken down into 1 of 6 route categories:

- **Class 1 Routes**: On-Street Striped - Part of the roadway and delineated by pavement markings and regulatory signage. The lane which can be shared with in-line skaters, is usually located next to curb lane parking, and may include a marked buffer zone.
- **Class 2 Routes**: On-Street Proposed Route - A proposed on-street bicycle route (no pavement markings) that is recommended for bicycle travel. These routes may become On-Street Striped routes at which point their classification will change.
- **Class 3 Routes**: Greenway Paths - Separated from the roadway and delineated by pavement markings and regulatory signage. Bicycle paths are usually shared with multiple users, including pedestrians, runners and skaters. These paths usually follow waterfronts, rail lines, highways and parks.

- **Class 4 Routes**: Greenway Connectors/Signed Route - Shared use of the roadway, typically designated with informational signs. These routes often connect to greenway paths, but may also serve as a connection between on-street bike lanes.
- **Class 5 Routes**: Planned/Proposed Greenways - A proposed bicycle path separated from the roadway, these paths usually follow waterfronts, rail lines, highways and parks. These routes may become Greenway Paths in the future at which point their classification will change.
- **Class 6 Routes**: Signed Route - Shared use of the roadway, typically designated with informational signs.

Using the original PDF bicycle route maps of NYC from 2002, bicycle routes were digitized and attributed appropriately using the above 6 route categories. Bicycle routes from 2007 were acquired in shapefile format from the NYD Department of City Planning and were already digitized and attributed with the above 6 route categories.

♠ Variables are available for 2002 and 2007, where "YY" in the variable name denotes the year.

Existing Bicycle Route Variables

BIKLENYYC1 – Total length in Km of **Class 1** bicycle routes per ZIP Code.
BIKPCTYYC1 – Total percentage of **Class 1** bicycle routes [BIKLENYYC1 / BIKLENYY_T].
BIKLENYYC3 – Total length in Km of **Class 3** bicycle routes per ZIP Code.
BIKPCTYYC3 – Total percentage of **Class 3** bicycle routes [BIKLENYYC3 / BIKLENYY_T].
BIKLENYYC4 – Total length in Km of **Class 4** bicycle routes per ZIP Code.
BIKPCTYYC4 – Total percentage of **Class 4** bicycle routes [BIKLENYYC4 / BIKLENYY_T].
BIKLENYYC6 – Total length in Km of **Class 6** bicycle routes per ZIP Code.
BIKPCTYYC6 – Total percentage of **Class 6** bicycle routes [BIKLENYYC6 / BIKLENYY_T].
BIKLENYY_A – Total length in Km of **Existing** bicycle routes per ZIP Code.
BIKPCTYY_A – Total percentage of **Existing** bicycle routes [BIKLENYY_A / BIKLENYY_T].
BIKDENYY_A – Density of **Existing** bicycle routes per Km2 [BIKLENYY_A / ZIP_LNDKM2].

Proposed or Planned Bicycle Route Variables

BIKLENYYC2 – Total length in Km of **Class 2** bicycle routes per ZIP Code.
BIKPCTYYC2 – Total percentage of **Class 2** bicycle routes [BIKLENYYC2 / BIKLENYY_A].
BIKLENYYC5 – Total length in Km of **Class 5** bicycle routes per ZIP Code.
BIKPCTYYC5 – Total percentage of **Class 5** bicycle routes [BIKLENYYC5 / BIKLENYY_A].
BIKLENYY_P – Total length in Km of **Proposed** or **planned** bicycle routes per ZIP Code.
BIKPCTYY_P – Total percentage of **Proposed** or **Planned** bicycle routes [BIKLENYY_P / BIKLENYY_T].
BIKDENYY_P – Density of **Proposed** or **Planned** bicycle routes per Km2 [BIKLENYY_P / ZIP_LNDKM2].

Existing, Proposed, or Planned Bicycle Route Variables

BIKLENYY_T – Total length in Km of **Existing**, **Proposed**, or **Planned** bicycle routes per ZIP Code.
BIKDENYY_T – Density of **Existing**, **Proposed**, or **Planned** bicycle routes per Km2 [BIKLENYY_T / ZIP_LNDZM2].

Dun and Bradstreet Variables: 2001 & 2005

Dun & Bradstreet <www.dnb.com>, a commercial vendor of business data, maintains a comprehensive database of micro-data on over 11 million U.S. business locations. This data was purchased for NYC and represents the food and retail presence on the ground. Variables are available for 2001 and 2005, where "Y" in the variable name denotes the last digit of the respective year.

Food Environment Categories: Point-in-Polygons Counts

DUNBRADTOY – Count of Dun & Bradstreet businesses per ZIP Code. Included SIC codes at the 4-digit level: **5411** – Grocery Stores • **5421** – Meat and Fish Markets • **5431** – Fruit and Vegetable Markets • **5441** – Candy, Nut, and Confectionery Store • **5451** – Dairy Products Stores • **5461** – Retail Bakeries • **5499** – Miscellaneous Food Stores • **5812** – Eating Places.

"PRIM_SIC" = '5411' OR "PRIM_SIC" = '5421' OR "PRIM_SIC" = '5431' OR "PRIM_SIC" = '5441' OR "PRIM_SIC" = '5451' OR "PRIM_SIC" = '5461' OR "PRIM_SIC" = '5499' OR "PRIM_SIC" = '5812'

EATPLC_TOY – Count of Dun & Bradstreet eating place businesses [SIC = 5812] per ZIP Code. "PRIM_SIC" = '5812'

EATPLCXTRY – EATPLC_TOT plus national chains, local chains, or pizza (selected by name) records that do not have the eating place SIC code 5812 (e.g. a Popeye's that classified itself as a drinking place).
"AFASTPIZZA" = 1 OR "PRIM_SIC" = '5812'

OTHERRESY – Restaurants NOT classified in one of the following fast food or pizza categories ("EATPLC_TOT" =1 and removed records where: "NATL_CHAIN" =1 OR "LOCFASTCHN" =1 OR "LOCSICFAST" =1 OR "LOCFASTFOD" =1 OR "ALLFASTFOD" =1 OR "PIZZA" =1 OR "NTLCHNPIZA" =1 OR "AFASTPIZZA" =1). These are all restaurants not classified as fast food by any of our possible definitions.

NATL_CHAIY: National chain restaurant variable [If TRADESTYLE or COMPNAME = ARBY'S, AU BON PAIN, BASKIN-ROBBINS, BEN & JERRY'S, BLIMPIE SUBS & SALADS, BOJANGLES, BOSTON MARKET, BURGER KING, CARVEL ICE CREAM CAKES, CHECKERS, CHIPOTLE, CHURCHS CHICKEN, COLD STONE CREAMERY, COSI, DAIRY QUEEN, DAYLIGHT DONUTS, DOMINO'S PIZZA, DUNKIN' DONUTS, EL POLLO LOCO, HAAGEN-DAZS, HARDEE'S, I CAN'T BELIEVE IT'S YOGURT, JAMBA JUICE, KFC, KRISPY KREME, LITTLE CAESARS PIZZA, LONG JOHN SILVER'S, MC DONALD'S, PANDA EXPRESS, PAPA JOHN'S PIZZA, PIZZA DELIGHT, PIZZA HUT, POPEYE'S, QUIZNOS, RED ROBIN, ROUND TABLE PIZZA, SBARRO, SCHLOTZSKY'S DELI, STARBUCKS, SUBWAY, TACO BELL, TCBY, TOGOS, WENDY'S, WHITE CASTLE].

This list is a combination of Catie's coding and the franchises included in InfoUSA's franchise codes. Only a few were added from the InfoUSA list because Catie's list had the majority.

See SPSS code for selection criteria (Code for Fast Food Variables). After applying the code, went through and removed the records where the SIC code was not an "eating place" or one of the other food SIC codes.
Selected NATL_CHAIN = 1, selected from selection the following:
"PRIM_SIC" = '5411' OR "PRIM_SIC" = '5421' OR "PRIM_SIC" = '5431' OR "PRIM_SIC" = '5441' OR "PRIM_SIC" = '5451' OR "PRIM_SIC" = '5461' OR "PRIM_SIC" = '5499' OR "PRIM_SIC" = '5812'
Switched the selection and then selected NATL_CHAIN=1 from this switched selection – marked these '0'. Did the same for LOCFASTCHN.

LOCFASTCHY: This variable identifies local fast food by name. NATL_CHAIN = 0 AND [IF TRADESTYLE OR COMP_NAME = Crown Fried Chicken, Kennedy Fried Chicken, Kosher Delight Corp, Miami Subs & Grill, El Pollo Supremo, Supreme Chicken Of New Jersey, Nathans Famous Inc, Pudgies Famous Chicken, Zorn Famous Chicken & Ribs, Grays Papaya, Manhattan Bagel, Roy Rogers, Everything Yogurt & Salad, Coffee Shop North Central Hosp, Directors Metro Food Service, Chicken Holiday, Auntie Annes, Metropolitan Deli, Xando, OR Papaya King].

LOCSICFASY: This variable identifies local fast food by SIC code—restaurants that self identified as fast food via their SIC coding.
By SIC coding: [IF (NATL_CHAIN = 0) AND SIC = 58120300, 58120307, 58120308, 58120302]
Query code: "PRIM8_EXT1" ='58120300' OR "PRIM8_EXT2" ='58120300' OR "PRIM8_EXT3" ='58120300' OR "PRIM8_EXT4" ='58120300' OR "PRIM8_EXT1" ='58120307' OR "PRIM8_EXT2" ='58120307' OR "PRIM8_EXT3" ='58120307' OR "PRIM8_EXT4" ='58120307' OR "PRIM8_EXT1" = '58120308' OR "PRIM8_EXT2" ='58120308' OR "PRIM8_EXT3" ='58120308' OR "PRIM8_EXT4" ='58120308' OR "PRIM8_EXT1" = ' 58120302' OR "PRIM8_EXT2" =' 58120302' OR "PRIM8_EXT3" =' 58120302' OR "PRIM8_EXT4" =' 58120302'

LOCFASTFOY: The combination of LOCFASTCHN and LOCSICFAST. Restaurants that were either a local fast food chain (by name) or self-identified as fast food. "LOCSICFAST" =1 OR "LOCFASTCHN" =1

ALLFASTFOY – Variable that puts national and local fast food into one variable.
"NATL_CHAIN" =1 OR "LOCFASTFOD" =1

PIZZAY – Pizza variable [IF (NATL_CHAIN = 0 and LOCFASTFOD=0) AND SIC = 58120600, 58120601, 58120602, OR TRADESTYLE = PIZZA OR COMP_NAME = PIZZA].

Query code: "PRIM8_EXT1" ='58120600' OR "PRIM8_EXT2" ='58120600' OR "PRIM8_EXT3" ='58120600' OR "PRIM8_EXT4" = '58120600' OR "PRIM8_EXT1" ='58120601' OR "PRIM8_EXT2" ='58120601' OR "PRIM8_EXT3" ='58120601' OR "PRIM8_EXT4" ='58120601' OR "PRIM8_EXT1" ='58120602' OR "PRIM8_EXT2" ='58120602' OR "PRIM8_EXT3" ='58120602' OR "PRIM8_EXT4" ='58120602' OR "COMP_NAME" LIKE '%PIZZA%' OR "TRADESTYLE" LIKE '%PIZZA%'

NTLCHNPIZY – National chain and pizza variable [OF ((NATL_CHAIN = 1) OR (PIZZA = 1))].

"NATL_CHAIN" =1 OR "PIZZA" =1

AFASTPIZZY – All fast food and pizza places [IF (ALLFASTFOD = 1 OR PIZZA = 1)].

"ALLFASTFOD" =1 OR "PIZZA" =1

SUPRMARKEY: All supermarkets where SIC = 5411, sales volume is >= $2,000,000 or total number of employees is >= 18.

Query code: "PRIM_SIC" ='5411' AND "SALES_VOL" >='000000002000000' –marked these.
"PRIM_SIC" = '5411' AND "SALES_VOL" = '000000000000000' AND "EMPL_HERE" >= '000000018' – marked these.

GROCERYY: All grocery stores where SIC = 5411, sales volume is < $2,000,000 and the total number of employees is < 18 and > 4. The selection is less than $2 mil AND between 17 and 5 employees because if you choose $2 mil OR between 17 and 5 employees, you will also capture the < 5 employee establishments, which is how bodegas have been defined.

Query code: "PRIM_SIC" ='5411' AND "SALES_VOL" < '000000002000000' AND "EMPL_HERE" < '000000018' AND "EMPL_HERE" > '000000004'.
- Also had to mark stores as grocery stores when their sales volumes were less than $2 mil (and not = $0) and their number of employees was equal to or greater than 18.
- "PRIM_SIC" = '5411' AND "SALES_VOL" < '000000002000000' AND "EMPL_HERE" >= '000000018' AND "SALES_VOL" <> '000000000000000'

BODEGASY: All bodegas where SIC = 5411 and the total number of employees is <= 4 and Grocery does not equal 1 and supermarket does not equal 1.

Query code: "PRIM_SIC" ='5411' AND "EMPL_HERE" <= '000000004' AND "SUPRMARKET" <>1 AND "GROCERY" <>1.

CONVENIENY: All conveniences stores where SIC = 541102 (in any of the 4 SIC fields). Convenience stores are mutually exclusive with supermarkets, grocery stores, and bodegas. Removed convenience stores from supermarkets, grocery, and bodegas, so places that code themselves as convenience stores are first convenience stores.

Query code: "PRIM8_EXT1" LIKE '541102%' OR "PRIM8_EXT2" LIKE '541102%' OR "PRIM8_EXT3" LIKE '541102%' OR "PRIM8_EXT4" LIKE '541102%'

FRUITVEGIY: Fruit and vegetable markets where SIC = 5431 ("PRIM_SIC" ='5431').

FISHY – Fish/seafood markets where SIC6_EXT1 = 542101 OR SIC6_EXT2 = 542101 OR SIC6_EXT3 = 542101 OR SIC6_EXT4 = 542101.

"PRIM8_EXT1" LIKE '542101%' OR "PRIM8_EXT2" LIKE '542101%' OR "PRIM8_EXT3" LIKE '542101%' OR "PRIM8_EXT4" LIKE '542101%'

MEATY – Meat markets where SIC6_EXT1 = 542102 OR SIC6_EXT2 = 542102 OR SIC6_EXT3 = 542102 OR SIC6_EXT4 = 542102.

"PRIM8_EXT1" LIKE '542102%' OR "PRIM8_EXT2" LIKE '542102%' OR "PRIM8_EXT3" LIKE '542102%' OR "PRIM8_EXT4" LIKE '542102%'

MEATORFISY: Meat or fish markets that are not classified beyond the 4-digit primary SIC code in addition to those that are (either Fish or Meat markets from above).

"PRIM_SIC" ='5421' AND "FISH" = 1 OR "MEAT" = 1

CANDYNUTY – Candy, nut, and confectionery stores where SIC = 5441.
"PRIM_SIC" ='5441'

BAKERYY: Retail bakeries where SIC = 5461.
"PRIM_SIC" ='5461' AND "NATL_CHAIN" <>1
also removed Pizza and Local Chain fast food (17 records) from the bakery category

NATURALFOY: Natural food stores including health and dietetic food stores [SIC = 549901].
"PRIM8_EXT1" LIKE '549901%' OR "PRIM8_EXT2" LIKE '549901%' OR "PRIM8_EXT3" LIKE '549901%' OR "PRIM8_EXT4" LIKE '549901%'

SPECIALTYY: Special stores including dairy products stores [SIC = 5451] and miscellaneous food stores [SIC = 5499] excluding health and dietetic food stores [SIC = 549901] and fast food or pizza.
"PRIM_SIC" ='5451' OR "PRIM_SIC" ='5499' AND "NATURALFOD" <>1
removed from the selection: "NATL_CHAIN" =1 OR "LOCFASTCHN" =1 OR "PIZZA" =1

DRINKINGY – Drinking places where SIC = 5813
"PRIM_SIC" ='5813'

Checked each variable and fixed overlaps using the following logic rules:

- National Chain or Pizza → National Chain
- Local fast food (either by name or SIC) or Pizza → Local fast food
- National Chain or Bakery → National Chain
- Local Chain or Bakery → Local Chain
- Pizza or Bakery → Pizza
- Bodega or Pizza → Bodega (same goes for other grocery stores or convenience stores)

HEALTHYY – Count of BMI-healthy food outlets (NYC Wide–2001: n=1,514; 2005: n=1,833). This category includes Supermarkets, Fruit and Veggie Markets, and Natural Food Stores ["SUPRMARKET" = 1 OR "FRUITVEGIE" = 1 OR "NATURALFOD" = 1].

UNHEALTHYY – Count of BMI-unhealthy food outlets (NYC Wide–2001: n=12,071; 2005: n=17,348). This category includes Fast Food Restaurants, Pizza Restaurants, Convenience Stores, Bodegas, Bakeries, Candy and Nut Stores, and Meat Markets ["BAKERY" = 1 OR "CANDYNUT" = 1 OR "MEAT" = 1 OR "CONVENIENC" = 1 OR "BODEGAS" = 1 OR "AFASTPIZZA" = 1].

NEUTRALY – Count of BMI-neutral or intermediate food outlets (NYC Wide–2001: n=15,780; 2005: n=20,671). This category includes Other Restaurants (non-Fast Food), Grocery Stores, Fish Markets, and Specialty Food Stores ["GROCERY" = 1 OR "FISH" = 1 OR "SPECIALTY" = 1 OR "OTHERREST" = 1].

Food Environment Categories: Kernel Density Measures

DBYY_H_AVG – Kernel Density Estimation variable: **average** number of BMI-healthy food outlets per Km2.
DBYY_H_MED – Kernel Density Estimation variable: **median** number of BMI-healthy food outlets per Km2.
DBYY_N_AVG – Kernel Density Estimation variable: **average** number of BMI-neutral food outlets per Km2.
DBYY_N_MED – Kernel Density Estimation variable: **median** number of BMI-neutral food outlets per Km2.
DBYY_U_AVG – Kernel Density Estimation variable: **average** number of BMI-unhealthy food outlets per Km2.
DBYY_U_MED – Kernel Density Estimation variable: **median** number of BMI-unhealthy food outlets per Km2.

Retail Environment Categories: Point-in-Polygons Counts

HARDWAREX – Hardware stores where SIC = 5251.
"PRIM_SIC" ='5251'

DRUG_STORX – Drug stores and proprietary stores where SIC = 5912.
"PRIM_SIC" ='5912'

BANKX – National commercial banks, state commercial banks, commercial banks (nec), federal savings institutions, and savings institutions (except federal) where SIC = 6021 OR SIC = 6022 OR SIC = 6029 OR SIC = 6035 OR SIC = 6036.
"PRIM_SIC" ='6021' OR "PRIM_SIC" ='6022' OR "PRIM_SIC" ='6029' OR "PRIM_SIC" ='6035' OR "PRIM_SIC" ='6036'

CREDIT_UX – Federal and state credit unions where SIC = 6061 OR SIC = 6062.
"PRIM_SIC" ='6061' OR "PRIM_SIC" ='6062'

CHILDCAREX – Child day care facilities where SIC = 8351.
"PRIM_SIC" ='8351'

LAUNDRYX – Coin operated laundries and cleaning facilities where SIC = 7215.
"PRIM_SIC" ='7215'
GROOMINGX – Beauty and barber shops where SIC = 7231 OR SIC = 7241.
"PRIM_SIC" ='7231' OR "PRIM_SIC" ='7241'

Sidewalk Café Variables

Sidewalk cafés were geocoded from a list of Legally Operating New York City Sidewalk Cafés (as of June 8, 2006) which was provided to BEH by the New York City Department of Consumer Affairs. For more information on Legally Operating Sidewalk Cafés, please visit http://home2.nyc.gov/html/dca/html/licenses/swc_main.shtml

CAFE06_CNT – Count of sidewalk cafés per ZIP Code as of 2006.
CAFE06_DEN – Density of sidewalk cafés per Km2 [CAFE06_CNT / ZIP_LNDKM2].
CAFE06_AVG – Kernel Density Estimation variable: **average** number of sidewalk cafés per Km2.
CAFE06_MED – Kernel Density Estimation variable: **median** number of sidewalk cafés per Km2.

Poultry and Small Animal Slaughterhouse Variables

A list of licensed Poultry and Small Animal Slaughterhouses were acquired from Jennifer Jensen (j_jensen@planning.nyc.gov) with the Division of Housing, Economic & Infrastructure Planning, Department of City Planning representing Slaughterhouses as of 2004 (a more specific date in 2004 is not known). This data was then cleaned and geocoded using GeoSupport software.

MEAT04_CNT – Count of slaughterhouses per ZIP Code as of 2004.
MEAT04_DEN – Density of slaughterhouses per Km2 [MEAT04_CNT / ZIP_LNDKM2].
MEAT04_AVG – Kernel Density Estimation variable: **average** number of slaughterhouses per Km2.
MEAT04_MED – Kernel Density Estimation variable: **median** number of slaughterhouses per Km2.

Farmers' Market Variables

Lists of Farmers' Markets were acquired and reconciled for 2006 from various sources, including:

- New York City Coalition Against Hunger (http://www.nyccah.org/)
- Council on the Environment of New York City (http://www.cenyc.org/)
- Farmers' Market Federation of New York (http://www.nyfarmersmarket.com/)

FMRK06_CNT – Count of farmers' markets per ZIP Code as of 2006.
FMRK06_DEN – Density of farmers' markets per Km2 [FMRK06_CNT / ZIP_LNDKM2].
FMRK06_AVG – Kernel Density Estimation variable: **average** number of farmers markets per Km2.
FMRK06_MED – Kernel Density Estimation variable: **median** number of farmers markets per Km2.

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.

ZIP_RESKM2 – Residential Building area in Km2 (ZIP_LUMIXA input variable).
ZIP_RESPCT – % Residential Building area [([ZIP_RESKM2] / ([ZIP_COMKM2] + [ZIP_RESKM2])) * 100].
ZIP_COMKM2 – Commercial Building area in Km2 (ZIP_LUMIXA input variable).
ZIP_COMPCT – % Commercial Building area [([ZIP_COMKM2] / ([ZIP_RESKM2] + [ZIP_COMKM2])) * 100].
ZIP_LUMIXA – Land Use Mix A. This measure was calculated using the PLUTO tax lot data available from the NYC Department of City Planning. A co-distribution of commercial and residential building area was derived from the PLUTO 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.

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

ZIP_LU01KM – One and Two Family Building area Km2.
ZIP_LU01PT – Percent of One and Two Family Building area [ZIP_LU01KM / ZIP_LNDKM2].
ZIP_LU02KM – Multi–Family Walk-up Building area in Km2.
ZIP_LU02PT – Percent of Multi–Family Walk-up Building area [ZIP_LU02KM / ZIP_LNDKM2].
ZIP_LU03KM – Multi–Family Elevator Building area in Km2.
ZIP_LU03PT – Percent of Multi–Family Elevator Building area [ZIP_LU03KM / ZIP_LNDKM2].
ZIP_LU04KM – Mixed Residential and Commercial Building area in Km2.
ZIP_LU04PT – Percent of Mixed Residential and Commercial Building area [ZIP_LU04KM / ZIP_LNDKM2].
ZIP_LU05KM – Commercial and Office Building area in Km2.
ZIP_LU05PT – Percent of Commercial and Office Building area [ZIP_LU05KM / ZIP_LNDKM2].
ZIP_LU06KM – Industrial and Manufacturing area in Km2.
ZIP_LU06PT – Percent of Industrial and Manufacturing area [ZIP_LU06KM / ZIP_LNDKM2].
ZIP_LU07KM – Transportation and Utility area in Km2.
ZIP_LU07PT – Percent of Transportation and Utility area [ZIP_LU07KM / ZIP_LNDKM2].

ZIP_LU08KM – Public Facilities and Institutions area in Km2.
ZIP_LU08PT – Percent of Public Facilities and Institutions area [ZIP_LU08KM / ZIP_LNDKM2].
ZIP_LU09KM – Open Space and Outdoor Recreation area in Km2.
ZIP_LU09PT – Percent of Open Space and Outdoor Recreation area [ZIP_LU09KM / ZIP_LNDKM2].
ZIP_LU10KM – Parking Facilities area in Km2.
ZIP_LU10PT – Percent of Parking Facilities area [ZIP_LU10KM / ZIP_LNDKM2].
ZIP_LU11KM – Vacant Land area in Km2.
ZIP_LU11PT – Percent of Vacant Land area [ZIP_LU11KM / ZIP_LNDKM2].
ZIP_LUORKM – Other area in Km2 (parcels not classified as category 1-11 but still a legal tax lot).
ZIP_LUORPT – Percent of Other area [ZIP_LUOTR / ZIP_LNDKM2].
ZIP_LUTOT – Total area of all 11 categories plus the Other Land Use category in Km2.
ZIP_LUPTOT – Percent of the 11 categories plus the Other Land Use category area [ZIP_LUTOT / ZIP_LNDKM2].
ZIP_ROWKM2 – Total Right-of-Way area in Km2 (areas not defined as tax lots, e.g., streets, sidewalks, etc.),
ZIP_ROWPCT – Percent Right-of-Way area [ZIP_AREAROW / ZIP_LNDKM2].

Street Tree Variables: 2005-2006

Street trees are thought to improve pedestrian comfort (shade, wind protection) and the attractiveness of the street. These are measures of the total number of street trees, density of street trees by ZIP Code area in Km2, density of street trees by total length of all streets per ZIP Code in Km, and the average and median number of street trees per ZIP Code per Km2 as derived from kernel density grid surfaces.

Point-level data from the City of New York, Parks & Recreation's 2005-06 Street Tree Census was used to create the metrics. To learn more, please visit: http://www.nycgovparks.org/sub_your_park/trees_greenstreets/treescount/.

TREE05_CNT – Count of street trees per ZIP Code as of 2005-06.
TREE05_DEN – Density of street trees per Km2 of land area [TREE05_CNT / ZIP_LNDKM2].
TREE05_STR – Density of street trees per Km of road length [TREE05_CNT / ZIP_STRKM].
TREE05_AVG – Kernel Density Estimation variable: **average** number of street trees per Km2.
TREE05_MED – Kernel Density Estimation variable: **median** number of street trees per Km2.

Homicide Crime Variables: 2003-2006

The metrics available here are of the total number of homicides, density of homicides by ZIP Code area in Km2, and the average and median number of homicides per ZIP Code per Km2 as derived from kernel density grid surfaces for individual years and averaged across all years.

The locations of these homicide variables were mined from this New York Times web-site: http://projects.nytimes.com/crime/homicides/map/

HOMC03_CNT – Count of homicides in 2003 per ZIP Code.
HOMC03_DEN – Density of homicides per Km2 [HOMC03_CNT / ZIP_LNDKM2].
HOMC03_AVG – Kernel Density Estimation variable: **average** number of homicides per Km2.
HOMC03_MED – Kernel Density Estimation variable: **median** number of homicides per Km2.

HOMC04_CNT – Count of homicides in 2004 per ZIP Code.
HOMC04_DEN – Density of homicides per Km2 [HOMC04_CNT / ZIP_LNDKM2].
HOMC04_AVG – Kernel Density Estimation variable: **average** number of homicides per Km2.
HOMC04_MED – Kernel Density Estimation variable: **median** number of homicides per Km2.

HOMC05_CNT – Count of homicides in 2005 per ZIP Code.
HOMC05_DEN – Density of homicides per Km2 [HOMC05_CNT / ZIP_LNDKM2].
HOMC05_AVG – Kernel Density Estimation variable: **average** number of homicides per Km2.
HOMC05_MED – Kernel Density Estimation variable: **median** number of homicides per Km2.

HOMC06_CNT – Count of homicides in 2006 per ZIP Code.
HOMC06_DEN – Density of homicides per Km2 [HOMC06_CNT / ZIP_LNDKM2].
HOMC06_AVG – Kernel Density Estimation variable: **average** number of homicides per Km2.
HOMC06_MED – Kernel Density Estimation variable: **median** number of homicides per Km2.

HOMCIDECNT – Count of homicides in 2003-06 per ZIP Code [HOMC03_CNT + HOMC04_CNT + HOMC05_CNT + HOMC06_CNT].
HOMCIDEDEN – Density of homicides per Km2 [(HOMC03_CNT + HOMC04_CNT + HOMC05_CNT + HOMC06_CNT) / ZIP_LNDKM2].
HOMCIDEAVG – Kernel Density Estimation variable: **average** number of homicides per Km2 [(HOMC03_AVG + HOMC04_AVG + HOMC05_AVG + HOMC06_AVG) / 4].
HOMCIDEMED – Kernel Density Estimation variable: **median** number of homicides per Km2.

Pedestrian–Bicyclist–Motorist Crash Variables

Crash data comes from the New York State Department of Transportation. Variables are aggregated by both pedestrian—motorist injuries, pedestrian—motorist fatalities, bicyclist—motorist injuries, and bicyclist—motorist fatalities. Variables are available for 2002 through 2006, where "YY" in the variable name denotes the year.

PINJYY_CNT – Count of pedestrian-motorist injuries per ZIP Code.
PINJYY_DEN – Density of pedestrian-motorist injuries per Km2 [PINJYY_CNT / ZIP_LNDKM2].
BINJYY_CNT – Count of bicyclist-motorist injuries per ZIP Code.
BINJYY_DEN – Density of bicyclist-motorist injuries per Km2 [BINJYY_CNT / ZIP_LNDKM2].
PKILYY_CNT – Count of pedestrian-motorist fatalities per ZIP Code.
PKILYY_DEN – Density of pedestrian-motorist fatalities per Km2 [PKILYY_CNT / ZIP_LNDKM2].
BKILYY_CNT – Count of bicyclist-motorist fatalities per ZIP Code.
BKILYY_DEN – Density of bicyclist-motorist fatalities per Km2 [BKILYY_CNT / ZIP_LNDKM2].

Department of Sanitation Sidewalk and Street Cleanliness Scorecard Variables

Project Scorecard Overview

Since 1973, the City of New York has used a litter measurement tool called "Project Scorecard" to evaluate and measure the cleanliness of its streets and sidewalks. Trained evaluation teams use the Scorecard to "rate" the degree of surface litter by comparing actual street conditions to photographic standards. This measurement tool provides objective information about street and sidewalk cleanliness. Project Scorecard is managed by the Mayor's Office of Operations because they are responsible for tracking and monitoring the City's cleanliness over time. To learn more about the purpose, measurement methods, etc. of the Project Scorecard, please visit: <http://www.usmayors.org/USCM/best_practices/litter/NewYork.html>.

Sidewalk and Street Scorecard Variables: Individual Years

Scorecard variables are aggregated to 1 of 234 NYC Sanitation Sections. Variables are available from 2002-2006, where "YY" in the variable name denotes the year. These variables were created using an area-weighted methodology as Sanitation Sections do not nest within ZIP Codes.

STCLNYYAVG – Mean street cleanliness as a numerical value ranging from 1.0 [cleanest] to 3.0 [filthiest]. Ratings below 1.5 are considered "acceptable" according to both the consensus arrived at in public surveys conducted by Scorecard and the Department's own operating standards.

STACCYYPCT – Percent streets acceptable as a range of 0.0 to 100.0.
STFILYYPCT – Percent streets filthy as a range of 0.0 to 100.0.
SWCLNYYAVG – Mean sidewalk cleanliness as a numerical value ranging from 1.0 [cleanest] to 3.0 [filthiest]. Ratings below 1.5 are considered "acceptable" according to both the consensus arrived at in public surveys conducted by Scorecard and the Department's own operating standards.
SWACCYYPCT – Percent sidewalks acceptable as a range of 0.0 to 100.0.
SWFILYYPCT – Percent sidewalks filthy as a range of 0.0 to 100.0.

Sidewalk and Street Scorecard Variables: Averaged Across Years

Scorecard variables were also averaged across 2002-2006.

STCLEANWGT – Mean street cleanliness as a numerical value ranging from 1.0 [cleanest] to 3.0 [filthiest]. Ratings below 1.5 are considered "acceptable" according to both the consensus arrived at in public surveys conducted by Scorecard and the Department's own operating standards.
STACCPTWGT – Percent streets acceptable as a range of 0.0 to 100.0.
STFILTYWGT – Percent streets filthy as a range of 0.0 to 100.0.
SWCLEANWGT – Mean sidewalk cleanliness as a numerical value ranging from 1.0 [cleanest] to 3.0 [filthiest]. Ratings below 1.5 are considered "acceptable" according to both the consensus arrived at in public surveys conducted by Scorecard and the Department's own operating standards.
SWACCPTWGT – Percent sidewalks acceptable as a range of 0.0 to 100.0.
SWFILTYWGT – Percent sidewalks filthy as a range of 0.0 to 100.0.

Walkability Index Variables



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.

To date, BEH has created and used two different versions of the "Walkability Index." These indexes were created by z-scoring and then summing and/or averaging various components of the built and social environments which are further discussed below. Mike Bader will provide a Stata .do file that documents the components used for each version and also performs the z-scoring and performs the construction of the two "Walkability Index" variables.

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.

Several of the variables employed in the "Walkability Index" were previously calculated and are aforementioned in this data dictionary. To avoid duplication of variables in the dataset itself, the "Walkability Index" variables flagged with a "♦" are also found in other sections of this data dictionary but are listed again here in an abridged definition for the sake of convenience. The variables are not, however, duplicated in the dataset.

♦ **ZIP_LNDKM2** – Total land area in Km2 (inland water features subtracted out).
♦ **ZIP_POPDEN** – Population density per Km2.
♦ **ZIP_STRINT** – Count of unique street intersections.

♦ **ZIP_INTDEN** – Density of unique streets intersections per Km2.
♦ **BUS04_CNT** – Count of unique MTA bus stops.
♦ **BUS04_DEN** – Density of unique bus stops per Km2.
♦ **SUBW07_CNT** – Count of MTA subway stations.
♦ **SUBW07_DEN** – Density of subway stations per Km2.
♦ **ZIP_COMPCT** – Percent Commercial Building area.
♦ **ZIP_LUMIXA** – Land Use Mix A. This measure was calculated using the PLUTO tax lot data available from the NYC Department of City Planning. A co-distribution of commercial and residential building area was derived from the PLUTO 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.

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

RETLFAR – Retail floor area ratio – Retail building floor area divided by retail land area, both measures drawn from the MapPLUTO database.
RESDEN1 – Density of residential units per residential building area – the number of residential units over the total residential building floor area measurement geography in Km2 [RES_UNITS / B3].
RESDEN2 – Density of residential units – the number of housing units divided by the total ZIP Code land area in Km2 [RES_UNITS / ZIP_LNDKM2].
ENTROPY – 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 PLUTO database. We used the PLUTO 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 represented in the census tract. 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 represented (between 1 and 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])

NUM_LOTS – Total number of unique PLUTO tax lots per CHS ZIP Code.
RES_UNITS – Total number of residential units in all the buildings per CHS ZIP Code.

Breezy Point Tax Lot Value Imputation Process

CHS ZIP Code 11697 lies at the western tip of the Rockaway peninsula, between Rockaway Inlet and Jamaica Bay. The neighborhood is governed by Queens Community Board 14. The community is run by the Breezy Point Cooperative, in which all residents pay the maintenance, security, and community-oriented costs involved with keeping the community private. The cooperative owns the entire 500-acre (2 Km2) community; residents own their homes and hold shares in the cooperative. Within this ZIP Code the majority of the buildings fall within two PLUTO tax lots which both also happen to fall within Breezy Point. However, the PLUTO data variables do not correctly report the number of Residential Units or the Residential and Commercial building areas. These 3 variables are required to calculate the Land Use Mix A and Walkability Index variables. Therefore, their values were imputed.

The process of imputation involved using a GIS layer of building footprints to determine the number of buildings per tax lot. A second GIS layer of building elevations was spatially associated with the building footprints so the elevation of each building could be determined. Next, using Google Street View several buildings were identified simultaneously in Google and in a GIS environment in order to determine an elevation cut-point between 1 and 2 story buildings. Once enough buildings had been identified and objectively compared, a cut-point of 29 feet was used to distinguish 1 and 2 story buildings. The tallest building was 40 feet which was also identified in Google Street View to verify it wasn't a 3 story building, which it was not, suggesting that only 1 and 2 story buildings exist. Then, the square footage of each building was calculated using the building footprints layer and multiplied by the number of stories to determine the total square footage of each building.

To determine commercial vs. residential buildings, Google Street View was again used in combination with prior *in situ* knowledge, and the NumBldgs, UnitsRes, and UnitsTotal PLUTO variables for one of the two tax lots (PLUTO_ID = 311514) as the values appeared to be correct. The values reported that there were 19 commercial buildings. And since each tax lot shares an equal length of road frontage along Rockaway Point Blvd, the road where all of the commercial businesses are believed to be located (based on *in situ* investigation) a value of 19 was attributed to both tax lots as the number of commercial buildings. Finally, building footprints were spatially associated with the two tax lots and the following variables aggregated to each:

- total number of buildings;
- total number of commercial buildings;
- total number of residential buildings;
- total square footage commercial buildings; and
- total square footage residential buildings.

The actual values assigned to each tax lot are as follows:

Tax Lot 1 (PLUTO_ID = 298197)

-NumBldgs: 1816	
-ResUnits: 1797	(1816 - 19)
-BldgArea: 3002833.474657 ft2	(AvgBldgsArea: 1653.542662)
-ResArea: 2971416.164079 ft2	(3002833.474657 - (1653.542662 * 19))
-ComArea: 31417.310578 ft2	(1653.542662 * 19)
-RetailArea: 31417.310578 ft2	(1653.542662 * 19)

Tax Lot 2 (PLUTO_ID = 311514)

-NumBldgs: 597	
-ResUnits: 578	(597 - 19)
-BldgArea: 1131305.939767 ft2	(AvgBldgsArea: 1894.984824)
-ResArea: 1095301.228111 ft2	(1131305.939767 - (1894.984824 * 19))
-ComArea: 36004.711656 ft2	(1894.984824 * 19)
-RetailArea: 36004.711656 ft2	(1894.984824 * 19)

Actual name of the MapPLUTO variables whose values were imputed:
-BldgArea
-ComArea
-ResArea
-RetailArea (same value as ComArea as it is assumed that retail area is the only type of commercial area present)
-NumBldgs
-UnitsRes
-UnitsTotal

ALR Digital Pedestrian Count: Number of People on the Street Estimate

The digital version of the pedestrian count variable was constructed using population data from the U.S. Census for block groups, commercial zoning data from the NYC Department of City Planning's PLUTO database, and subway stop and ridership data from the New York State Department of Transportation and the NYC Transit Authority. These variables were included in this measure because the number of people walking on the street is a factor of the density of the population living in the area, the density of the commercial attractions, and the population passing through on public transportation.

Population density was calculated for all of NYC from block group centroids (geometric center points) using kernel density—with a 100 foot cell size and a search radius of one mile, or 5,280 feet. The kernel density function transforms population values from a series of points to a continuous raster surface with grid cells containing the values for population per square mile. Similarly, commercial zoning density was calculated for all of NYC from tax lot centroids, also using kernel density, a 100 foot cell size and a one mile search radius. The commercially zoned tax lots were first isolated to create a raster grid surface that represented the density of tax lots zoned for commercial uses. Last, subway ridership density was calculated for NYC from subway stop points and ridership values that are associated with these points. Again, the kernel density transformation used a 100 foot cell size and a one mile search radius.

The cell values for each of these three raster grids were then averaged (separately) by the areas of the sampled block polygons. The average values were then transformed to be on scale similar to the pedestrian counts that observers recorded on the street. Population density average values were divided by 10,000, commercial zoning density average values were divided by 100 and subway ridership density average values were divided by 10,000. Last, the three transformed average values were summed for the final digital pedestrian count estimate [ZIP_NUMPEP].

ZIP_POPTOT – Transformed population density average per Mi2 [POPTOT_AVG / 10,000].
ZIP_COMLOT – Transformed commercial zoning density average per Mi2 [COMLOT_AVG / 100].
ZIP_SUBPOP – Transformed population density average per Mi2 [SUBPOP_AVG / 10,000].
ZIP_NUMPEP – Composite ALR digital pedestrian count estimate per Mi2 [ZIP_POPTOT + ZIP_COMLOT + ZIP_SUBPOP].

NYC Park Inspection Program

The Park Inspection Program (PIP) is a comprehensive, outcome-based performance measurement system that generates frequent, random, and detailed inspections of the conditions and quality of the City's public parks, playgrounds, and recreational facilities. Trained inspectors use hand-held computers and digital cameras to conduct nearly 5,000 inspections of ~3,000 properties per year and visit each park once or twice per year. Parks larger than 6 acres are divided into zones and rated separately; playgrounds are also rated separately. The two Figures below uses Central Park of an example of how large parks are divided into zones for PIP rating purposes.

The Table below lists the sixteen features which are split into three categories and which are rated as "Acceptable" or "Unacceptable" during field inspections. The four cleanliness features (litter, glass, weeds, and graffiti) are rated at every park and playground and are used to determine each park's Cleanliness rating of each park or playground. The other twelve features are only rated when available. For example, when rating a park with no amenities only the four Cleanliness features can be rated. However, if a playground is being rated seven or more features are rated, the primary which are: litter, glass, weeds, graffiti, paved surfaces, play equipment, and safety surfaces. All sixteen features are used to determine the Overall Condition of a park. The **CLEANLINESS** condition is rated unacceptable if any two cleanliness features are unacceptable or any one cleanliness feature is very unacceptable. **OVERALL** condition is rated unacceptable if the site fails cleanliness, if any three of any of its features are unacceptable, or if any one of its features is very unacceptable.

To learn more about PIP, please visit: http://www.nycgovparks.org/sub_about/parks_numbers/pip.html.

Figure. – Central Park in Manhattan and its PIP rated zones.

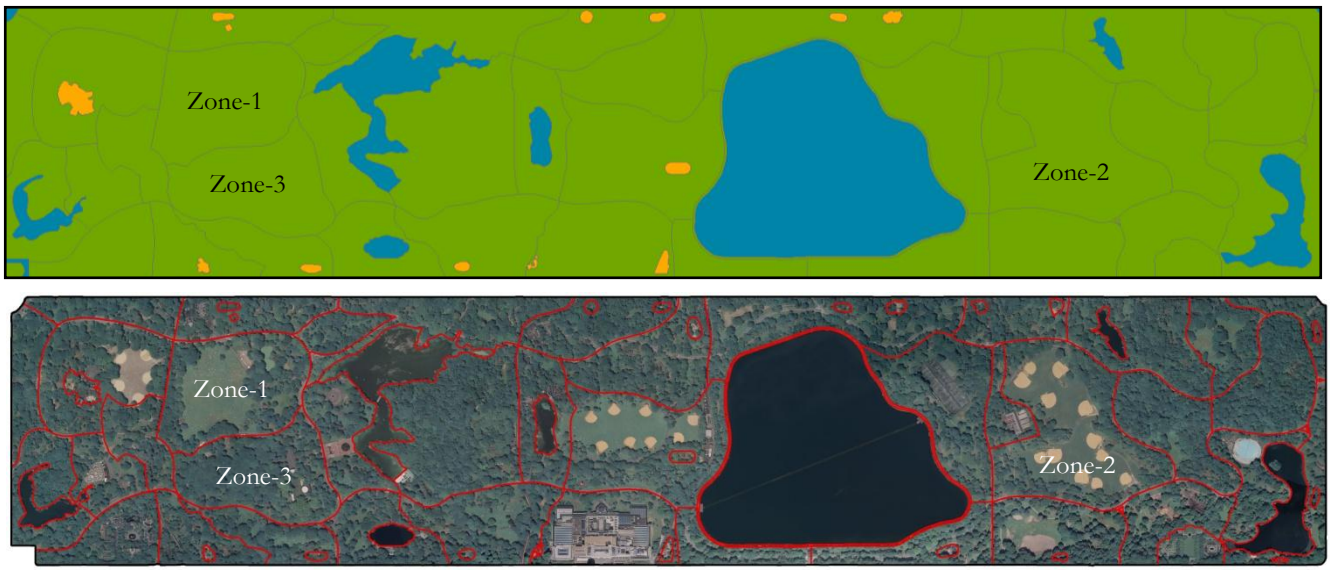


Table. – Sixteen PIP rated park and playground features.

Cleanliness

- Litter
- Glass
- Weeds
- Graffiti

Landscape

- Lawns
- Trees
- Athletic Fields
- Horticultural Areas
- Water Bodies
- Trails

Structural

- Sidewalks
- Paved Surfaces
- Benches
- Fences
- Play Equipment
- Safety Surfaces

Park Inspection Program [PIP] Variables

♠ Variables consider **ALL** park features regardless of the Category or Subcategory Type, **NOTHING** was removed, not Greenstreets not traffic triangle parks, not sitting area park features; nothing.
ALLPARKCNT – Count of **ALL** park features per ZIP Code.
ALLPARKPCT – Percent of ZIP Code occupied by **ALL** park features [ALLPARKKM2 / ZIP_LNDKM2].
ALLPARKKM2 – Land area in Km2 of **ALL** park features per ZIP Code.

♠ Variables consider a reduced number **SELECT** park features, selected based on their Category or Subcategory Type. Park features removed include: "CATMASTER" = ' ' OR "CATMASTER" = 'Greenstreet' OR "CATMASTER" = 'Other' OR "CATMASTER" = 'Water' OR "SCATMASTER" = ' ' OR "SCATMASTER" = 'Cemetery' OR "SCATMASTER" = 'DOT Adopt-A-Highway' OR "SCATMASTER" = 'Golf Course' OR "SCATMASTER" = 'Greenstreet' OR "SCATMASTER" = 'Greenthumb Garden' OR "SCATMASTER" = 'Greenthumb' OR "SCATMASTER" = 'Highway Property' OR "SCATMASTER" = 'Island' OR "SCATMASTER" = 'Natural Area' OR "SCATMASTER" = 'Park Strip' OR "SCATMASTER" = 'Parking Lot' OR "SCATMASTER" = 'Pier' OR "SCATMASTER" = 'Private Property' OR "SCATMASTER" = 'Sitting Area/Triangle/Mall' OR "SCATMASTER" = 'To Be Determined' OR "SCATMASTER" = 'Type 1' OR "SCATMASTER" = 'Type 2' OR "SCATMASTER" = 'Type 3' OR "SCATMASTER" = 'Undeveloped Parkland'. **THIS INCLUDES ALL PARK FEATURES, THAT AVOIDED EXCLUSION FROM THE ABOVE QUERY, WHETHER THEY WERE PIP RATED OR NOT**. [Query Expression: REMOVE_JQ = 1].
SELPARKCNT – Count of **SELECT** park features per ZIP Code.
SELPARKPCT – Percent of ZIP Code occupied by **SELECT** park features [SELPARKKM2 / ZIP_LNDKM2].
SELPARKKM2 – Land area in Km2 of **SELECT** park features per ZIP Code.

♠ Variables consider **ONLY** the PIP rated playgrounds features.
PLGDPIPCNT – Count of **ONLY** the PIP rated playgrounds per ZIP Code.
PLGDPIPPCT – Percent of ZIP Code occupied by **ONLY** PIP rated playgrounds [PLGDPIPKM2 / ZIP_LNDKM2].
PLGDPIPKM2 – Land area in Km2 of **ONLY** the PIP rated playgrounds per ZIP Code.

Count, Area, and Percentage Variables

♠ Variables consider only those park features whose Category Type is Greenstreet, Small Park, Large Park, or Other, regardless of whether they were PIP rated or not PIP rated.
ALL_PRKCNT – Count of all parks [ALL_GRNCNT + ALL_SMLCNT + ALL_LRGCNT + ALL_OTRCNT].
ALL_PRKPCT – Percent of ZIP Code that is covered by all parks.
ALL_PRKKM2 – Land area in Km2 of ZIP Code that is covered by all parks.
ALL_GRNCNT – Count of all greenstreets.
ALL_GRNPCT – Percent of ZIP Code that is covered by all greenstreets.
ALL_GRNKM2 – Land area in Km2 of ZIP Code that is covered by all greenstreets.
ALL_SMLCNT – Count of all small parks [< 6 acres].
ALL_SMLPCT – Percent of ZIP Code that is covered by all small parks.
ALL_SMLKM2 – Land area in Km2 of ZIP Code that is covered by all small parks.
ALL_LRGCNT – Count of all large parks [>= 6 acres].
ALL_LRGPCT – Percent of ZIP Code that is covered by all large parks.
ALL_LRGKM2 – Land area in Km2 of ZIP Code that is covered by all large parks.
ALL_OTRCNT – Count of all "other" park features [not categorized as greenstreets, small or large parks].
ALL_OTRPCT – Percent of ZIP Code that is covered by all "other" park features.
ALL_OTRKM2 – Land area in Km2 of ZIP Code that is covered by all "other" park features.

♠ Variables consider only those park features whose Category Type is Greenstreet, Small Park, Large Park, or Other, AND were PIP rated one or more times between 2002-2006.
PIP_PRKCNT – Count of PIP rated parks [PIP_GRNCNT + PIP_SMLCNT + PIP_LRGCNT + PIP_OTRCNT].
PIP_PRKPCT – Percent of ZIP Code that is covered by PIP rated parks.
PIP_PRKKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated parks.
PIP_GRNCNT – Count of PIP rated greenstreets.
PIP_GRNPCT – Percent of ZIP Code that is covered by PIP rated greenstreets.
PIP_GRNKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated greenstreets.
PIP_SMLCNT – Count of PIP rated small parks [< 6 acres].

PIP_SMLPCT – Percent of ZIP Code that is covered by PIP rated small parks.
PIP_SMLKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated small parks.
PIP_LRGCNT – Count of PIP rated large parks [>= 6 acres].
PIP_LRGPCT – Percent of ZIP Code that is covered by PIP rated large parks.
PIP_LRGKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated large parks.
PIP_OTRCNT – Count of PIP rated "other" park features [not categorized as greenstreets, small or large parks].
PIP_OTRPCT – Percent of ZIP Code that is covered by PIP rated "other" park features.
PIP_OTRKM2 – Land area in Km2 of ZIP Code that is covered by PIP rated "other" park features.

♠ Variable considers only those **park** and **playground** features whose Category Type is Small Park or Large Park AND if they were PIP rated one or more times between 2002-2006.
PARK_PIP – Count of PIP rated park and playground features per ZIP Code used to calculate the PIP rated variables. This count may differ from values found in PIP_PRKCNT because only those park features whose Category Type is Small Park or Large Park are used for PIP variable related analysis and the PIP_PRKCNT variable includes Greenstreets, Small Parks, Large Parks, and Other categories.

♠ The below count values (i.e., PARKPIP_02 - PARKPIP_06) represent, out of the Count available in PARK_PIP, how many times were those total available park and playground features were PIP rated in each year.
PARKPIP_02 – Number of times the available park and playground features were PIP rated per ZIP Code in 2002.
PARKPIP_03 – Number of times the available park and playground features were PIP rated per ZIP Code in 2003.
PARKPIP_04 – Number of times the available park and playground features were PIP rated per ZIP Code in 2004.
PARKPIP_05 – Number of times the available park and playground features were PIP rated per ZIP Code in 2005.
PARKPIP_06 – Number of times the available park and playground features were PIP rated per ZIP Code in 2006.

♠ Variable considers only those park features whose Subcategory Type is DOE Playgrounds, Jointly Operated Playgrounds, Neighborhood Playgrounds, and Playgrounds within Parks…meaning, all **playground** features only ("SCATMASTER" = 'DOE Plgd' OR "SCATMASTER" = 'Jointly Operated Playground' OR "SCATMASTER" = 'Neighborhood Playground' OR "SCATMASTER" = 'Playground Within Park') AND if they were PIP rated one or more times between 2002-2006.
PLGD_PIP – Count of PIP rated playground features per ZIP Code used to calculate the PIP rated variables. This count is the same as those values found in PIP_PRKCNT but is listed again here for the purpose of contiguity.

♠ The below count values (i.e., PLGDPIP_02 - PLGDPIP_06) represent, out of the total count available in PLGD_PIP, how many times were those total available playground features were PIP rated in each year.
PLGDPIP_02 – Number of times the available playground features were PIP rated per ZIP Code in 2002.
PLGDPIP_03 – Number of times the available playground features were PIP rated per ZIP Code in 2003.
PLGDPIP_04 – Number of times the available playground features were PIP rated per ZIP Code in 2004.
PLGDPIP_05 – Number of times the available playground features were PIP rated per ZIP Code in 2005.
PLGDPIP_06 – Number of times the available playground features were PIP rated per ZIP Code in 2006.

Average Park and Playground Rating Score Variables: Individual Years

Using the raw PIPs data from 2002-2006, the seven available park and playground feature inspection ratings (weeds, glass, litter, graffiti, play equipment, safety surfaces and paved areas) were first coded 0 for satisfactory and 1 for unsatisfactory. Next, because park and playground features, can, and usually are, inspected two or more times per year, an average score for each playground feature was calculated across inspections done in a given year (i.e., scale range is still 0-1). Then, CHS merged ZIP Codes were intersected with the park and playgrounds (PIP rated only) in order to identify which park and playground features fell partially or completely within each ZIP Code. The seven available playground ratings were then summed by each unique ZIP and by each year. Finally, the seven PIP ratings that had just been summed by ZIP and year were again averaged for each ZIP and each year, where the numerator was the summed value of each rating and the dominator was the number of times the park and playground features had been rated. The results are variables representing the average PIP quality score for each of the seven measures available for park and playground features.

HIGHER SCORES INDICATE THAT THE PROPERTY DID WORSE; FOR THE SEVEN **INDIVIDUAL** MEASURES, THE AVERAGE SCORE HAVE A POSSIBLE RANGE OF 0-1; THE COMBINED **CLEANLINESS** MEASURE HAS A POSSIBLE RANGE OF 0-4; AND THE COMBINED **QUALITY** MEASURE HAS A POSSIBLE RANGE OF 0-6.

The below variables were calculated using all available PIP rated park and playground features per ZIP Code and are available for 2002-2006, where "YY" in the variable name denotes the year.

WEEDSYYAVG – Average weeds score.
LITERYYAVG – Average litter score.
GLASSYYAVG – Average glass score.
GRAFIYYAVG – Average graffiti score.
PLYEQYYAVG – Average play equipment score.
SAFTYYYAVG – Average safety surfaces score.
PAVDSYYAVG – Average paved surfaces score.
CLNSRYYAVG – Average combined cleanliness score.
SUMSRYYAVG – Average combined quality score.

Average Park and Playground Rating Score Variables: Averaged Across Years

After the seven available park (weeds, glass, litter, and graffiti) and playground feature inspection ratings (weeds, glass, litter, graffiti, play equipment, safety surfaces and paved areas) had been averaged individually by ZIP and year, an average for each of the seven ratings was calculated and averaged across all years. This was accomplished by simply summing the seven ratings for 2002-2006 by ZIP, and then dividing by 5.

HIGHER SCORES INDICATE THAT THE PROPERTY DID WORSE; FOR THE SEVEN **INDIVIDUAL** MEASURES, THE AVERAGE SCORE HAVE A POSSIBLE RANGE OF 0-1; THE COMBINED **CLEANLINESS** MEASURE HAS A POSSIBLE RANGE OF 0-4; AND THE COMBINED **QUALITY** MEASURE HAS A POSSIBLE RANGE OF 0-6.

The below variables were calculated using all available PIP rated park and playground features per ZIP Code.

WEEDS_AVG – Average weeds score.
LITTER_AVG – Average litter score.
GLASS_AVG – Average glass score.
GRAFIT_AVG – Average graffiti score.
PLAYEQ_AVG – Average play equipment score.
SAFETY_AVG – Average safety surfaces score.
PAVEDS_AVG – Average paved surfaces score.
CLNSCR_AVG – Average combined cleanliness score.
SUMSCR_AVG – Average combined quality score.

Minimum–Maximum Park and Playground Rating Score Variables: Individual Years

The minimum and maximum PIP quality scores were calculated for each of the four PIP measures available for parks (weeds, glass, litter, and graffiti) and the seven PIP measures available for playground features (weeds, glass, litter, graffiti, play equipment, paved surfaces, and safety surfaces).

HIGHER SCORES INDICATE THAT THE PROPERTY DID WORSE; FOR THE SEVEN **INDIVIDUAL** MEASURES, THE MINIMUM AND MAXIMUM SCORES HAVE A POSSIBLE RANGE OF 0-1; THE COMBINED **CLEANLINESS** MEASURE HAS A POSSIBLE RANGE OF 0-4; AND THE COMBINED **QUALITY** MEASURE HAS A POSSIBLE RANGE OF 0-6.

The below variables were calculated using all available PIP rated park and playground features per ZIP Code and are available for 2002-2006, where "YY" in the variable name denotes the year.

MINWEEDSYY – Minimum weeds score.
MAXWEEDSYY – Maximum weeds score.
MINLITERYY – Minimum litter score.
MAX LITERYY – Maximum litter score.
MINGLASSYY – Minimum glass score.
MAXGLASSYY – Maximum glass score.
MINGRAFIYY – Minimum graffiti score.
MAX GRAFIYY – Maximum graffiti score.
MINPLYEQYY – Minimum play equipment score.
MAXPLYEQYY – Maximum play equipment score.
MINSAFTYYY – Minimum safety surfaces score.
MAXSAFTYYY – Maximum safety surfaces score.
MINPAVDSYY – Minimum paved surfaces score.
MAXPAVDSYY – Maximum paved surfaces score.
MINCLNSRYY – Minimum combined cleanliness score.
MAXCLNSRYY – Maximum combined cleanliness score.
MINSUMSRYY – Minimum combined quality score.
MAXSUMSRYY – Maximum combined quality score.

Minimum–Maximum Park and Playground Rating Score Variables: Averaged Across Years

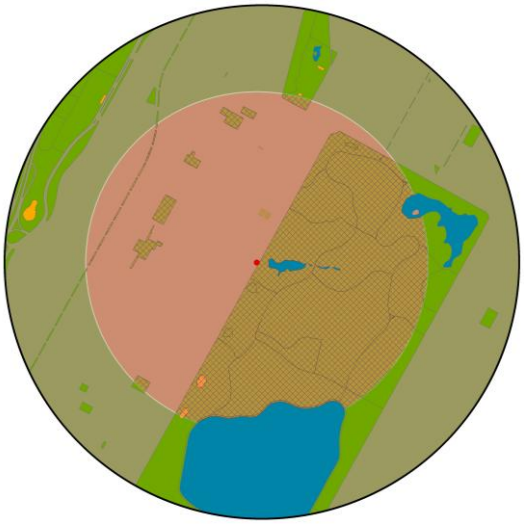
After the minimum and maximum PIP quality scores were calculated for each of the four PIP measures available for parks (weeds, glass, litter, and graffiti) and the seven PIP measures available for playground features (weeds, glass, litter, graffiti, play equipment, paved surfaces, and safety surfaces) individually by ZIP and year, the average for each of the seven ratings was calculated and averaged across all years. This was accomplished by simply summing the seven minimum and maximum ratings for 2002-2006 by ZIP, and then dividing by 5.

HIGHER SCORES INDICATE THAT THE PROPERTY DID WORSE; FOR THE SEVEN **INDIVIDUAL** MEASURES, THE MINIMUM AND MAXIMUM SCORES HAVE A POSSIBLE RANGE OF 0-1; THE COMBINED **CLEANLINESS** MEASURE HAS A POSSIBLE RANGE OF 0-4; AND THE COMBINED **QUALITY** MEASURE HAS A POSSIBLE RANGE OF 0-6.

The below variables were calculated using all available PIP rated park and playground features per ZIP Code.

MIN_WEEDS – Averaged minimum weeds score.
MAX_WEEDS – Averaged maximum weeds score.
MIN_LITTER – Averaged minimum litter score.
MAX_LITTER – Averaged maximum litter score.
MIN_GLASS – Averaged minimum glass score.
MAX_GLASS – Averaged maximum glass score.
MIN_GRAFFI – Averaged minimum graffiti score.
MAX_GRAFFI – Averaged maximum graffiti score.
MIN_PLAYEQ – Averaged minimum play equipment score.
MAX_PLAYEQ – Averaged maximum play equipment score.
MIN_SAFETY – Averaged minimum safety surfaces score.
MAX_SAFETY – Averaged maximum safety surfaces score.
MIN_PAVEDS – Averaged minimum paved surfaces score.
MAX_PAVEDS – Averaged maximum paved surfaces score.
MIN_CSCOR – Averaged minimum combined cleanliness score.
MAX_CSCOR – Averaged maximum combined cleanliness score.
MIN_SSCOR – Averaged minimum combined quality score.
MAX_SSCOR – Averaged maximum combined quality score.

Area-Weighted PIP Cleanliness Score Variables



Using PIPs data from 2002-2006, park inspection features (weeds, glass, litter, graffiti, play equipment, paved surfaces, and safety surfaces) were coded 0 for satisfactory and 1 for unsatisfactory. Within each property, the average score for each feature was calculated across inspections done in 2002-2006 and then averaged across all years. The scale has a range of 0-1 with higher scores being worse. For each property a total cleanliness score and a total quality score was calculated for each year as the sum of the average scores for weeds, glass, litter, graffiti, and again averaged. Higher scores indicate that the property did worse; individual cleanliness scores have a scale ranging from 0-1; the total cleanliness score has a scale ranging from 0-4; and the combined quality score has a scale ranging from 0-6.

Area-Weighted Method – Area-weighted PIP cleanliness score that considers the total area of the park and playground features that fell within each ZIP Code. This PIP area-weighted score considers only the portion of the park you can reach within each ZIP Code and does not factor in the size of the rest of the park outside if it were to across the geographic bounds of two or more ZIP Codes. Therefore, only the portion of a park that is actually within the ZIP Code is considered.

$$WGHT2\_PIP = \frac{\sum_{i=1}^{N} A_i \times R_i \times PIP_i}{\sum_{i=1}^{N} A_i \times R_i}$$

Where: the N parks in the buffer are indexed 1,2,…,N
Ai= total park area of park i
Ri = percentage of park i in buffer
PIPi= PIP cleanliness score of park i

WT2_WEEDS – Method 2: Area-weighted PIP rated weeds score.
WT2_LITTER – Method 2: Area-weighted PIP rated litter score.
WT2_GLASS – Method 2: Area-weighted PIP rated glass score.
WT2_GRAFFI – Method 2: Area-weighted PIP rated graffiti score.
WT2_PLAYEQ – Method 2: Area-weighted PIP rated play equipment score.
WT2_SAFETY – Method 2: Area-weighted PIP rated safety surfaces score.
WT2_PAVEDS – Method 2: Area-weighted PIP rated paved surfaces score.
WT2_CSCOR – Method 2: Area-weighted PIP rated combined cleanliness score.
WT2_SSCOR – Method 2: Area-weighted PIP rated combined quality score.

Park Athletic and Recreational Amenities

Facility counts are per ZIP Code. Athletic and recreational amenities are represented by either point or polygon features. For a point or polygon feature to be counted, it has to simply interest or touch any part of the ZIP Code.

The geographic features used to create these metrics (with exception of access points and playgrounds) were acquired from the NYC Data Mine web-site (www.nyc.gov/data/). Access point and playground data came from the BEH spatial data repository and are modified versions of datasets acquired from Parks and Recreation.

ACCESSPNTS – [points] – Count of park and playground access points.
BASEBALL – [polygons] – Count of baseball fields.
BASKETBALL – [polygons] – Count of basketball courts.
BATHROOMS – [polygons] – Count of open bathrooms.
BEACHZONES – [polygons] – Count of unique beach zones. Beach zones are administrative units designated by the Department of Parks and Recreation. Beaches are split into manageable sizes for the purpose of PIP rating. There are 181 unique beach zones in the City and 7 unique beaches.
BEACHES – [polygons] – Count of beaches.
CALLBOXES – [points] – Count of emergency call boxes.
FLAGPOLES – [points] – Count of flag poles.
GOLFCOURSE – [polygons] – Count of golf courses.
GRNTMBGRDN – [polygons] – Count of green-thumb gardens.
HANDBALL – [polygons] – Count of handball courts.
HOCKEY – [polygons] – Count of hockey rinks.
HOOPS – [points] – Count of basketball hoops (actual point locations of backboard and rim).
MONUMENTS – [points] – Count of monuments.
MULTIFIELD – [polygons] – Count of multi-purpose courts and fields (i.e., several different sports can be played).
PARKINGLOT – [polygons] – Count of park related parking lots.
PICNICAREA – [points] – Count of picnic areas.
PLAYGROUND – [polygons] – Count of playgrounds (includes DOE playgrounds (n=35), jointly operated playgrounds (n=265), neighborhood playgrounds (n=466), and playgrounds within parks (n=269)).
POOLS – [polygons] – Count of open pools.
SOCERFBALL – [polygons] – Count of soccer or "American" football fields.
SHOWERS – [points] – Count of active spray showers.
TENNIS – [polygons] – Count of tennis courts.
TRACKS – [polygons] – Count of running tracks.
VOLLEYBALL – [polygons] – Count of volleyball courts.
WATERFOUNT – [points] – Count of active water fountains.

References

Gattrell, A.C. et al. (1996). Spatial point pattern analysis and its application in geographical epidemiology, *Transactions of the Institute of British Geographers* 21: pp. 256–274.

Guagliardo, M.F. et al. (2004). Physician accessibility: an urban case study of pediatric providers, *Health & Place* 10: pp. 273–283.

Longley, P.A. et al. (2001) *Geographic Information Systems and Science.* Chichester, NY: John Wiley and Sons, Inc.

O'Sullivan, D. and D. Unwin. (2002). *Geographic information analysis.* Hoboken, NJ: John Wiley and Sons, Inc.

Scott, D.W. (1998). *Multivariate Density Estimation*. John Wiley and Sons, Inc.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

Wand, M.P. and M.C. Jones. (1995). *Kernel Smoothing.* London, UK: Chapman and Hall.