Methods: Part 1

For Part 1, Focus on Gentrification, some analyses use methods that differ from those used for the indicators presented in Part 2 or for the indicators listed at the city, borough, and community-district levels in Part 3.

US Census Bureau Data

In order to track changes over time, we use the Neighborhood Change Database (NCDB) 2010, which is compiled by GeoLytics and the Urban Institute with support from the Rockefeller Foundation (2010). The NCDB provides data from the US Census Bureau at the census tract level back to 1970 but recalculated to match the census tract boundaries from 2010. We supplement NCDB data with tract-level, five-year estimates from the 2005-2009 and 2010-2014 American Community Survey (ACS). Because the ACS 2005-2009 five-year tract level data is reported in 2000 census tract boundaries, we use census tract relationship files[1] provided by the US Census Bureau to recalculate the 2005–2009 estimates in 2010 census tract boundaries. The ACS 2010-2014 five-year estimates are already tabulated in 2010 census tract boundaries.

The combination of NCDB and ACS data allows us to examine current geographies over the course of several decades, and therefore to better understand how neighborhoods as they are defined today changed over time.

In general, in this year's focus chapter we look at neighborhoods based on average household income in 1990 and the rate of rent increases since then. We group neighborhoods into a set of *gentrifying*, *non-gentrifying*, and *higher-income* neighborhoods. While, throughout this report, we refer to both community districts and sub-borough areas (SBAs) as "neighborhoods," in the focus on gentrification section we exclusively use SBAs as our definition of "neighborhoods." For example, while we report crime rates at the community district level in Part 3 of this report, the crime rates used in the focus chapter are calculated at the SBA level (see below for further discussion of crime statistics presented in the focus chapter). For a more detailed definition of SBAs and other geographies, please refer to the main Methods section below.

The US Census Bureau does not regularly report data in summary tables for SBAs before 2000, but using the NCDB we can aggregate data for census tracts, defined consistently over time, up to the SBA level. Since SBAs are composed of whole census tracts, there is no weighting required for this aggregation. However, aggregating from the tract level does mean that we have to use average (mean) income, rather than median income, and average gross rent, rather than median gross rent.[2] We define indicators used in the focus section, where they differ from those used in other parts of this report, below.

When presenting shares and averages by neighborhood type (gentrifying, non-gentrifying, and higher-income), we aggregate from the SBA level, weighting by SBA population. For example, we calculate the share of the population with a college degree in gentrifying neighborhoods as the average of the share with a college degree in each gentrifying SBA, weighted by the population of each SBA.

Average Gross Rent

We first calculate the average gross rent (which includes rent plus certain utilities; see the main Methods section for more detail) at the census tract level. To calculate the average rent at the SBA level, we average across census tracts, weighting by census tract population. The decennial censuses in 1990 and 2000 surveyed rents differently from the way the ACS surveys rents; specifically, the censuses did not record rents for renter-occupied single-family houses on at least 10 acres of land. Because of this, the US Census Bureau advises that rent estimates from the ACS should be compared to rent estimates from the 2000 or prior decennial censuses with caution. However, since there are very few if any renter-occupied single-family homes on at least 10 acres of land in New York City, we do make such comparisons.

Sources: Neighborhood Change Database (1990, 2000), American Community Survey (2005-2009, 2010-2014), NYU Furman Center

1 https://www.census.gov/geo/maps-data/data/relationship.html

2 According to the ACS 2014 1-year estimate, average household income in New York City was $85,356 while median household income was $52,996; median gross rent was about $1,275 for New York City households in 2014, while average gross rent was about $1,300.

Average Household Income

We calculate the average household income at the census tract level. We then average across census tracts, weighting by census tract population, to derive the SBA's average household income.

Sources: Neighborhood Change Database (1990, 2000), American Community Survey (2005-2009, 2010-2014), NYU Furman Center

Population

Within each SBA, we add up the total population across census tracts. Total population estimates for all years are derived from the decennial census, as reported in the NCDB. However, when presenting shares of the population with a certain characteristic (such as the share with a college degree), the denominator is derived from the same source as the numerator.

Sources: Neighborhood Change Database (1990, 2000, 2010), NYU Furman Center

Housing Units

Within each SBA, we add up the total number of housing units across census tracts. The total count of housing units is derived, for all years, from the decennial census data, as reported in the NCDB.

Sources: Neighborhood Change Database (1990, 2000, 2010), NYU Furman Center

Share of Households that are Non-Family

This indicator measures nonfamily households as a share of total households. The census defines nonfamily households as single-person households and householders living only with non-relatives. We aggregate from the census tract level to the SBA, weighting by tract population.

Sources: Neighborhood Change Database (1990, 2000), American Community Survey (2005-2009, 2010-2014), NYU Furman Center

Share of Families with Children

This indicator measures the total number of families and subfamilies with children as a share of all families and subfamilies. The US Census Bureau defines a family as a householder living in a unit with one or more people who are related by birth, marriage, or adoption.

Sources: Neighborhood Change Database (1990, 2000), American Community Survey (2005-2009, 2010-2014), NYU Furman Center

Population by Race and Ethnicity

This indicator measures the Hispanic, non-Hispanic Asian, non-Hispanic black, and non-Hispanic white populations, each as a share of the total population. We do not present shares for those who identified as other races or as two or more races, but these populations are counted in the total population. As a result, the calculated shares do not sum to 100 percent. Shares were calculated at the tract level and then aggregated to the SBA level, weighting by tract population. All statistics on race and ethnicity, including those from 2010, are derived from decennial censuses, as reported in the NCDB, and not from the ACS.

Sources: Neighborhood Change Database (1990, 2000, 2010), NYU Furman Center

Number and Share of Persons below the Poverty Line

This indicator measures the population below the poverty level.[3] For 1990 and 2000, the poverty status is based on the income in the previous calendar year (that is, 1989 and 1999, respectively). For 2010–2014 five-year estimates from the ACS, poverty status is based on the income in the 12 months prior to the date that the respondent was interviewed, which could have been at any point during the five-year range. In either case, the denominator is the total population for whom poverty status was determined. We calculate this at the census tract level and sum up to the SBA level, weighting by tract population.

Sources: Neighborhood Change Database (1990, 2000), American Community Survey (2010-2014), NYU Furman Center

Share of Population Aged 25 and Older with a College Degree

This indicator measures the total population aged 25 and older holding a bachelor's degree or higher, as a share of the total population aged 25 and older. We calculate the share at the census tract level and aggregate to the SBA level, weighting by tract population.

Sources: Neighborhood Change Database (1990, 2000), American Community Survey (2005-2009, 2010-2014), NYU Furman Center

3 Poverty thresholds are determined by family size. For example, in 2014, a 4-person family including two related children under the age of 18 was $24,008 (see http://www.census.gov/hhes/www/poverty/data/threshld/index.html).

Share of Population Aged 20 to 34

This indicator measures the share of the total population that is at least 20 and no more than 34 years old. We calculate the share at the census tract level and then aggregate to the SBA level, weighting by tract population.

Sources: Neighborhood Change Database (1990, 2000), American Community Survey (2005-2009, 2010-2014), NYU Furman Center

Public Use Microdata Sample Indicators

In addition to the indicators listed above, which are derived from the NCDB supplemented by five-year estimates at the census tract level from the ACS summary tables, we examine some characteristics using the Public Use Microdata Samples (PUMS) from the 2000 decennial census as well as the ACS five-year estimates for 2005-2009 and 2010-2014. See the main Methods section for more information about PUMS data. Since the PUMS data allow us to look at individual (anonymized) households, we can look at households at different income levels and identify households that recently moved.

The definition of *recent mover* is slightly different for the 2000 census and the ACS. For the 2000 census, all respondents were interviewed in April of 2000, and we define a recent mover as a household who moved into their unit in 1999 or later (that is, within about 15 months prior to their interview). For the ACS, respondents are interviewed throughout the year (and, for five-year samples, throughout the five-year span), and we define recent movers as households who moved into their unit within the 12 months prior to the date of their interview. Thus, recent movers in the 2005-2009 five-year ACS sample include households interviewed in 2005 who had moved into their unit in 2004, as well as households interviewed in 2009 who had moved into their unit in 2008, but not, for example, households interviewed in 2006 who had moved into their unit in 2004.

In order to preserve the anonymity of the PUMS data, the US Census Bureau does not release precise information about where a household moved from.[4] We can only observe the SBA where the respondent lived when they were interviewed and (roughly) the county (or, in New York City, the borough) where they moved from. Since boroughs are bigger than SBAs, we cannot distinguish between households who moved, for example, within a single gentrifying neighborhood, into a gentrifying neighborhood from another gentrifying neighborhood, or into a gentrifying neighborhood from a non-gentrifying or higher-income neighborhood.

We use PUMS data from the census, ACS, or both to calculate the following indicators.

Age Composition of Recent Movers

This indicator measures the number of residents who recently moved and were aged 20 to 34, 45 to 54, or 55 and over. We provide both the total count in each age group and the count in each age group as a share of all recent movers.

Sources: American Community Survey (2010-2014), NYU Furman Center

Educational Attainment of Recent Movers

This indicator measures the number of residents aged 25 and older in each neighborhood type who moved recently and either have or do not have a bachelor's degree or higher. For each level of educational attainment, we provide the total count of residents in that group and the count as a share of all recent movers aged 25 and over.

Sources: American Community Survey (2010-2014), NYU Furman Center

Rent Burden Share (all renter households and by income level)

This indicator measures the share of households that are rent-burdened, including moderately and severely rent-burdened households. For more information about how rent burden is defined, see "Moderately Rent-Burdened Households" and "Severely Rent-Burdened Households" in the Indicator Definitions and Rankings sections. We present statistics for all renters in each neighborhood type, as well as for renters in different income bands, expressed as a percentage of the Area Median Income (AMI) as defined by the US Department of Housing and Urban Development (HUD). See "US Department of Housing and Urban Development Income and Rent Limits" in the main Methods section below for more information on the AMI figures. In all cases, we use the fiscal year 2014 HUD AMI definitions, and households are placed into income bands based on their income after it has been adjusted for inflation to be comparable.

Sources: US Census (2000), American Community Survey (2005-2009 and 2010-2014), NYU Furman Center (ACS PUMS)

4 Technically, we know the migration Public Use Microdata Area (PUMA) where the respondent moved from. In some cases, a migration PUMA may include several counties, but in the New York City metro area, migration PUMAs generally coincide with counties.

Affordability of Recently Available Units

In order to look at the affordability of units on the rental market, we look at the share of units that were recently available and were affordable to appropriately sized households at various income levels, expressed as percentages of AMI. The HUD AMI levels are based on household size, so in order to calculate affordability we need to determine appropriately sized households, which we define as households with one more member than there are bedrooms in the unit. So, for a studio (with no bedrooms), a single-person household is appropriately sized. For a one-bedroom unit, a two-person household is appropriately sized, and so on. For units with three or more bedrooms, we use the income levels for four-person households.

For each income level, we define a unit as affordable if the gross rent was no more than 30 percent of the monthly income. The unit was recently available if the occupant at the time of the interview had recently moved into the unit. (See "Public Use Microdata Sample Indicators" above for the definition of recent movers.)

Sources: US Census (2000), American Community Survey (2010-2014), NYU Furman Center

Crowding (all renter households and by income level)

This indicator measures the share of households who have more than one person for every room in the unit. Severe crowding indicates that there were more than 1.5 people per room in the unit. Kitchens are counted as rooms, but bathrooms, porches, balconies, foyers, and halls are not. The wording of the ACS questionnaire changed slightly in 2008 to clarify that a room need not be separated by a door (for example, they can be separated by archways) and that unfinished basements should not be counted. For this reason, and because of other changes to question wording, the US Census Bureau advises caution in comparing indicators derived from the count of rooms from before 2008 to those in 2008 or after[5]; the 2005-2009 five-year data include responses to both pre-2008 and post-2008 questionnaires.

As with rent burden (see above), we break crowding rates out by different income levels, based on the HUD AMI for New York City. See "US Department of Housing and Urban Development Income and Rent Limits" in the main Methods section below for more information on the AMI figures.

Sources: American Community Survey (2005-2009 and 2010-2014), NYU Furman Center

Crime Data

The crime data used in the focus section differs from that used in Part 2 and reported at the city, borough, and community-district level in Part 3. First, in order to look at crime rates going further back in time, we use data that the New York City Police Department (NYPD) reported to the Federal Bureau of Investigation (FBI) in compliance with the Uniform Crime Reporting (UCR) program; outside of the focus section, we use NYPD data on serious felonies classified according to the New York State Penal Law, which only goes back to 2003.

The UCR defines seven index crimes: "criminal homicide" (including murder and non-negligent manslaughter), "forcible rape" (which until 2011 excluded statutory rape offenses or any offenses where the victim was not female), "robbery," "aggravated assault," "burglary" (breaking or entering), "larceny-theft"[6] (excluding motor vehicle theft), and "motor vehicle theft." As with the serious felony data, we break the UCR index crimes into two subtypes: violent crime, which includes criminal homicide, forcible rape, robbery and aggravated assault, and property crime, which includes burglary, larceny-theft, and motor vehicle theft.

Unlike in the data on serious felonies as defined by the New York Penal Law, not all index crimes in the UCR data are necessarily felonies. Furthermore, the NYPD categorizes all rape offenses as serious felonies according to the New York Penal Law definition, while the UCR data does not. See below for more discussion of the serious felony data used outside of the focus section.

5 Furthermore, before 2008 all units with more than nine rooms were top-coded as "nine or more." Very few units in New York City (about 2.1% in 2014), however, have more than nine rooms.

6 The FBI also tracks arson, although arson is not tabulated in the NYPD's UCR data.

The other main departure from the way crime rates are computed in the focus section of this year's report is that crime rates are calculated at the sub-borough area level, rather than the community-district level, as in the rest of the report. We use a similar housing-unit weighting algorithm (described in the main methods section, with respect to community districts) to allocate crimes, which are reported at the precinct level, to sub-borough areas; precinct-level data are weighted to sub-borough areas using housing unit counts from the Department of City Planning's 2007 PLUTO database.

Whereas the crime rates in the rest of the report use population counts from the prior decennial census (that is, 2000 census population counts for crime rates in 2003 through 2009, and 2010 census population counts for crime rates in 2010 or later), in the focus chapter we use a linear extrapolation to estimate population for each year of crime data. We use 1990 census population counts, 2000 census population counts, American Community Survey (ACS) 2005-2009 five-year estimates (which we treat as 2007 data for the purposes of extrapolation), and ACS 2010-2014 five-year estimates (which we treat as 2012 data for the purposes of extrapolation).

Housing Court Filings

In order to look at indicators of evictions, we present data on the number of housing court cases for non-payment of rent, per 1,000 rental units. We limit filings to those for residential addresses. The count of rental units for this indicator comes from one-year estimates at the SBA level from the ACS pre-tabulated summary data. For 2003 and 2004, we use the count of rental units from 2005.

Sources: New York State Office of Court Administration (2003 to 2014), American Community Survey (2005 to 2014), NYU Furman Center

Methods: Parts 2 and 3

Geographic Definitions

This report presents information for the entire City of New York, for each of the five boroughs, and for the neighborhoods within each borough. The city defines neighborhoods by dividing the boroughs into 59 community districts (CDs); the US Census Bureau, however, divides the boroughs into 55 sub-borough areas (SBAs). This report provides data for community districts where available but otherwise employs data at the sub-borough level. The term *neighborhood* is used in this report to refer to both community districts and sub-borough areas even though they are larger than what many consider to be neighborhoods. We have included reference maps for community districts and sub-borough areas following this chapter.

Borough

New York City consists of five boroughs: the Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Each borough is represented by a borough president, an elected official who advises the mayor on issues related to his or her borough and, along with the borough board, makes recommendations concerning land use and the allocation of public services. Each borough is also a county. Counties are legal entities with boundaries defined by state law.

Community District (CD)

Community districts are political units unique to New York City. Each of the 59 community districts has a community board. The community board's members are appointed by the borough president, though at least half must be nominees of City Council members who represent the community district. The community boards review applications for zoning changes and other land use proposals and make recommendations for budget priorities. Each community board is assigned a number within its borough. The borough and this number uniquely identify each of the 59 community districts. Therefore, we designate each community district with a two-letter borough code and a two-digit community board code. For example, BK 02 is the community district represented by Brooklyn Community Board 2.

Sub-Borough Area (SBA)

Sub-borough areas are geographic units created by the US Census Bureau for the administration of the New York City Housing and Vacancy Survey and were designed to have similar boundaries to those of community districts. These same areas are also defined by the US Census Bureau as Public Use Microdata Areas (PUMAs), so we are able to use the two terms interchangeably. Sub-borough areas are referred to using a three-digit number, where the first digit signifies the borough (we number boroughs in alphabetical order, with the Bronx being one and Staten Island being five). There are 59 community districts in New York City but only 55 sub-borough areas. The US Census Bureau combined four pairs of community districts in creating the sub-borough areas to improve sampling and protect the confidentiality of respondents. These pairs are Mott Haven/Melrose (BX 01) and Hunts Point/Longwood (BX 02) in the Bronx (combined into SBA 101), Morrisania/Crotona (BX 03) and Belmont/East Tremont (BX 06) in the Bronx (combined into SBA 102), the Financial District (MN 01) and Greenwich Village/Soho (MN 02) in Manhattan (combined into SBA 301), and Clinton/Chelsea (MN 04) and Midtown (MN 05) in Manhattan (combined into SBA 303). Because sub-borough areas are constructed from Census tracts, their boundaries do not coincide precisely with community district boundaries, which generally follow major streets. However, they are similar enough that we use them interchangeably throughout this report. The US Census Bureau periodically updates its geographic boundaries for each decennial census, and so the shapes of sub-borough areas changed slightly between the 2011 and 2012 releases of the American Community Survey. Although we treat these different vintages of sub-borough areas as being consistent over time, we advise some caution when comparing estimates from 2014 to earlier years.

Rankings

This report includes rankings of the five boroughs and all 59 community districts or 55 sub-borough areas for each indicator. The neighborhood ranked first has the highest number or percentage for the measure, even if lower values of measure are considered "better" (such as with crime rates). When possible, we rank all 59 community districts, though we present ranks for the 55 sub-borough areas for those indicators—including all indicators drawn from US Census Bureau and Home Mortgage Disclosure Act sources—that aggregate easily to the sub-borough area level. In addition, a few indicators are not available for all neighborhoods,

so we provide rankings for a subset of neighborhoods. For instance, we report the median sale price per unit at the community district level for only the predominant housing type in that district. Therefore, for each housing type, we present rankings only for the subset of community districts where that housing type predominates.

Visualization in Geographic Information Systems

Maps displaying New York City-specific administrative and political boundaries use base map data provided by the New York City Department of City Planning's Bytes of the Big Apple program. These boundaries include boroughs, community districts, zoning boundaries, public streets, police precincts, school districts, and individual properties. Maps displaying data in geographic areas defined by the US Census Bureau—such as sub-borough areas, tracts, and ZIP-code tabulation areas—use base map data from Census TIGER products.

United States Census Sources

A number of the indicators presented in the *State of New York City's Housing and Neighborhoods* are derived from five data sources collected by the US Census Bureau. These sources are described below along with a discussion of issues of comparability across sources.

Decennial Census (Census)

From 1970 through 2000, the decennial census consisted of two parts: the "short form" that collected information from every person and about every housing unit in the country, and the "long form" that asked additional questions of a sample of people and households. The short form collected information on age, race, Hispanic or Latino origin, household relationship, sex, tenure, and vacancy status. The long form provided more in-depth information about personal and housing characteristics such as income, employment status, and housing costs. In this edition of the *State of New York City's Housing and Neighborhoods*, we use data from the decennial census short and long forms to derive demographic, economic, and housing measures for the years 1970, 1980, 1990, and 2000. To create most of these indicators, we use summary census data reported at the city, borough, and sub-borough area levels. In 2010, the decennial census only included the short form since most of the data that had previously been included in the long form were now reported in the American Community Survey (see below). While much of the decennial census short-form data is also found in the American Community Survey (such as the count of households), the two sources often report differing numbers for statistical and methodological reasons.

American Community Survey (ACS)

The American Community Survey is an annual survey that collects data similar to those formerly collected by the census long form described above. As with the long form, the ACS covers only a sample of individuals and housing units. However, the ACS uses a smaller sample: the long form covered one out of every six housing units while the ACS only covers one in 40 housing units each year. The US Census Bureau began developing the ACS in 1996, but reliable annual estimates for geographic areas with a population of 65,000 or more only became available in 2005. In December 2008, the US Census Bureau began releasing three-year rolling estimates for all geographic areas with populations of 20,000 or more. In December 2010, the US Census Bureau began releasing five-year rolling estimates for geographic areas as small as block groups. In 2015, when releasing data for the 2014 ACS, the three-year estimates were discontinued and in this report only one-year or five-year estimates are used. Multiyear estimates are referred to by the whole range of years covered (for example, 2010–2014) and should be interpreted as a measure of the conditions during the whole range; due to space constraints, however, multiyear estimates presented in tables in Part 3 are, where noted, labeled using only the final year of the range (that is, an indicator from the 2010–2014 ACS is listed under the heading "2014"). Most of the indicators from the ACS in this edition are derived from pre-compiled summary tables reported by the US Census Bureau for the city as a whole, individual boroughs, and PUMAs, which, as discussed above, are identical to New York City's sub-borough areas (and which are often referred to in this report as "neighborhoods").

For most city- and borough-level indicators, we report figures derived from one-year estimates from the ACS. However, for some indicators, due to the small sample size, one-year estimates can be prone to volatility and sampling error, which can make it difficult to reliably discern whether an

indicator's change from one year to the next represents a real change or a statistical anomaly. In order to reduce this uncertainty and draw valid conclusions from differences over both time and space, for select indicators we use five-year ACS estimates. Please see the Sampling section below for recommendations about making comparisons over time and across geographic levels.

Public Use Microdata Samples (PUMS)

While most decennial census- and ACS-derived indicators use pre-tabulated summary data that are reported at a given geography, we calculate some indicators by aggregating person- and household-level data to the desired geographic level. The US Census Bureau makes individual-level data available in Public Use Microdata Samples (PUMS), which are anonymized extracts from the confidential microdata that the US Census Bureau uses in its own calculations for the decennial census and the ACS. We use PUMS data to calculate the household income distribution, income diversity ratio, median rent for recent movers and by number of bedrooms, severe rent burden (low-income renters), moderate rent burden (low-income renters), and several indicators by racial and ethnic group in the New York City section of Part 3. The only geographic areas that ACS PUMS data identify for a household are its state and PUMA. New York City's PUMAs are completely coterminous with its city boundaries.

Comparisons Between US Census Bureau Products

The US Census Bureau makes continual adjustments to the decennial census and the ACS to improve the coverage of the surveys and accuracy of the results. These adjustments often make cross-year comparisons difficult. Below is a discussion of the key areas where changes in sampling, question construction, or other methods might affect the comparability of indicators that we report in the *State of New York City's Housing and Neighborhoods* over time. More information about comparability between US Census Bureau data sources is available at: https://www.census.gov/programs-surveys/acs/guidance/comparing-acs-data.html.

Sampling

Because the ACS is a sample survey, not a census, all indicators derived from it are estimates, not exact counts.[1] The ACS sample includes approximately three million housing units nationwide, including about 66,000 in New York City. Readers should treat all estimates with some skepticism and be aware that the true value may differ from the reported estimate. This is especially important when comparing small year-to-year changes in sample-derived estimates or with estimates that are derived from a reduced sample. For example, the median rent does not use the entire sample but just the subset of respondents who are renters. The median rent for recent movers draws on an even smaller sample.

Comparisons Between Different Sampling Intervals

In order to report more reliable estimates of ACS-derived indicators for smaller geographies (such as sub-borough areas) or small populations (such as people aged 16 to 19 for the disconnected youth indicator), we use multiyear ACS estimates. The US Census Bureau recommends using one-year estimates for areas with populations of at least 65,000; all sub-borough areas have populations that are above 100,000, but in some cases certain subsamples (for example, recent movers or low-income renters) are considerably smaller. Five-year estimates reflect data from five full years of surveys, allowing for much more robust and accurate estimates at the expense of being less current. Multiyear estimates should be interpreted as describing the conditions that existed during the full sample range, and therefore should not be compared directly to one-year estimates for any of the individual years in the range. For example, the rental vacancy rate in SBA 201 (Greenpoint/Williamsburg in Brooklyn) was 2.2% according to the 2010–2014 ACS. In Brooklyn as a whole, the rental vacancy rate was 3.1% according to the 2014 ACS. Since the estimate for SBA 201 is for the entire period from 2010 through 2014, it is not strictly comparable to the borough-wide number, which comes from 2014 alone; if the vacancy rate in Greenpoint/Williamsburg and in Brooklyn as a whole declined substantially between 2010–2014, the estimate for SBA 201 would include the higher vacancy rate in 2010 as well as the lower vacancy rate in 2014, while the borough-wide estimate would only use data from after the

1 Censuses have their own methodological problems, of course, and may systemically under- or over-count certain populations.

decrease. (And, if the vacancy rate increased in the interim, vice versa.) It is appropriate, however, to compare multiyear estimates to estimates for a single year that falls outside the multiyear range. For example, one could compare the 2010–2014 estimate to the 2006 estimate, since 2006 is not within the range of 2010–2014.

Multiyear estimates can be compared to other multiyear estimates of the same duration as long as the ranges do not overlap. So, the 2010–2014 estimates for one sub-borough area can be compared to the 2005–2009 estimates for that sub-borough area and to the 2010–2014 estimates for other sub-borough areas. To compare a neighborhood's multiyear ACS estimate to the rest of the city, it is more effective to use its ranking than to compare its multiyear neighborhood estimate to the city's single-year estimate.

Income and Rent

Question construction and data collection for income information differs between the decennial census and the ACS. The 1990 census asked for the respondent's 1989 income, and similarly the 2000 census asked for the respondent's 1999 income; thus incomes reported in 1990 and 2000 are all for one fixed period of time (calendar years 1989 and 1999 respectively). In contrast, the ACS asks for the respondent's income over the "past 12 months." As the US Census Bureau collects ACS responses on an ongoing basis throughout the year, these estimates are not directly comparable; for example, a 2014 ACS respondent who was interviewed in January 2014 would report income that was mostly earned in 2013, while a respondent who was interviewed in December 2014 would report income that was mostly earned in 2014. The US Census Bureau notes that a comparison study of the 2000 census and the 2000 ACS found that incomes reported in the census were about four percent higher than the incomes reported in the ACS. Because of the data collection methods mentioned above, adjacent years of ACS data may not have reference months in common; thus comparisons of income data between adjacent ACS years (for example, 2010 and 2011) should not be interpreted as precise comparisons of economic conditions in those years. The indicators that draw on the ACS income data include the income diversity ratio (from PUMS data), median household income, poverty rate, and poverty rate by age. As a result, year-to-year changes in these indicators should be interpreted with caution.

Except where otherwise noted, we adjust all dollar figures for inflation (to constant 2015 dollars) from the nominal dollar values reported by the US Census Bureau (see below for more on how we adjust for inflation). However, such nominal dollar values are generated by the US Census Bureau using different methods depending on the source of the data. For ACS estimates that are included in the pre-tabulated summary data, the US Census Bureau reports dollar amounts that have been inflated to the annual average for the survey year (for example, calendar year 2014 for the 2014 ACS) based on the monthly Consumer Price Index (CPI). Thus, respondents' incomes (and rents) are adjusted to account for the fact that some are interviewed early in the year and others are interviewed later in the year. Such an adjustment, however, may not fully account for changes in the state of the economy over the course of the year. For example, if unemployment were higher in 2013 than in 2014 respondents interviewed in January 2014 would be more likely to report zero earnings in the last twelve months than similar respondents interviewed in December 2014, independent of the price level in the economy as measured by the CPI. In order to ensure the anonymity of individual responses in the PUMS data, however, the US Census Bureau does not adjust each respondent's income (or rent) for inflation based upon the month in which they were interviewed; instead, the identical adjustment is applied for all respondents, whether they were interviewed early or late in the year. If the rate of inflation changed over the course of the year, the dollar figures from PUMS could be biased. Since rent and income are recorded at the same time, the moderate and severe rent burden for low-income renters, which are also calculated from PUMS data, should not exhibit this bias.

The decennial censuses in 1990 and 2000 surveyed rents differently from the way the ACS surveys rents; specifically, the censuses did not record rents for renter-occupied single-family houses on at least 10 acres of land. Because of this, the US Census Bureau advises that rent estimates from the ACS should be compared to rent estimates from the 2000 or prior decennial censuses with caution. However, since there are very few if any renter-occupied single-family homes on at least 10 acres of land in New York City, we do make such comparisons.

Indicator Notes

US Department of Housing and Urban Development Income and Rent Limits

The US Department of Housing and Urban Development (HUD) defines income eligibility limits for its Section 8 and HOME programs based on the area median income (AMI) in a metropolitan area. HUD determines three general income limits at 30, 50, and 80 percent of AMI for various household sizes. HUD does not publish income guidelines for households with more than eight members, although its methodology allows for their calculation. To ease computation, we apply the eight-person limits to these larger households. As of fiscal year 2015, HUD assigned category names to ranges of the area median income:

- *Extremely low-income* households fall **at or below 30 percent** of AMI
- *Very low-income* household*s* have incomes **above 30 and at or below 50 percent** of AMI
- *Low-income* households have incomes **above 50 and at or below 80 percent** of AMI (although this report uses "low-income" as shorthand for any household earning at or below the 80 percent limit, which described 63.5% of renter households and 35.2% of owner-occupied households in New York City in 2014)

We employ HUD's general method to calculate 120 and 165 percent of the area median income for various household sizes. While HUD does not set category names for higher income ranges, we define *moderate-income* households as

Table 1: HUD Income Limits and Maximum Affordable Rents for New York City, 2014

	Extremely Low-Income	Very Low-Income	Low-Income	Low-Income	Moderate-Income	Moderate-Income	Middle-Income
Percentage of HUD Area Median Income	30%	50%	60%	80%	100%	130%	165%
Number of People in Household			**Income Limits (Nominal 2014$)**				
1	$17,650	$29,400	$35,250	$47,000	$58,750	$76,350	$96,900
2	$20,150	$33,600	$40,250	$53,700	$67,100	$87,250	$110,750
3	$22,650	$37,800	$45,300	$60,400	$75,500	$98,150	$124,600
4	$25,150	$41,950	$50,350	$67,100	$83,900	$109,050	$138,450
5	$27,900	$45,350	$54,350	$72,500	$90,600	$117,800	$149,500
6	$31,950	$48,700	$58,400	$77,850	$97,300	$126,500	$160,600
7	$36,050	$52,050	$62,400	$83,250	$104,050	$135,250	$171,650
8	$40,100	$55,400	$66,450	$88,600	$110,750	$143,950	$182,750
			Maximum Affordable Rent (Nominal 2014$)				
1	$441	$735	$881	$1,175	$1,469	$1,909	$2,423
2	$504	$840	$1,006	$1,343	$1,678	$2,181	$2,769
3	$566	$945	$1,133	$1,510	$1,888	$2,454	$3,115
4	$629	$1,049	$1,259	$1,678	$2,098	$2,726	$3,461
5	$698	$1,134	$1,359	$1,813	$2,265	$2,945	$3,738
6	$799	$1,218	$1,460	$1,946	$2,433	$3,163	$4,015
7	$901	$1,301	$1,560	$2,081	$2,601	$3,381	$4,291
8	$1,003	$1,385	$1,661	$2,215	$2,769	$3,599	$4,569

those making more than 80 and up to 120 percent of AMI, and *middle-income* households as earning more than 120 and up to 165 percent of AMI. Table 1 displays these income limits in nominal terms by household size for fiscal year 2014, along with the concomitant maximum affordable rents, which are calculated as 30 percent of the income limits.[2] For more information about HUD's method and their published guidelines, refer to individual years' guidelines at http://www.huduser.org/portal/datasets/il.html.

In order to calculate the share of rental units that are affordable to households of various income levels, we need to take household size into account, since the definition of income limits (and thus maximum affordable housing costs) vary by household size. For a rental unit with "n" bedrooms, we classify it as affordable at "X" percent of AMI if its gross rent is less than the maximum affordable rent specified by HUD for a household of size "n"+1; that is, a studio (i.e. a unit with zero bedrooms) is classified according to the maximum rent values for single-person households, a one-bedroom is classified according to the maximum rent values for two-person households, a two-bedroom is classified according to the maximum rent values for three-person households, and a unit with three or more bedrooms is classified according to the maximum rent values for four-person households. This method makes assumptions about the composition of the households that occupy each unit. Therefore, this indicator should be interpreted with some caution.

Index of Housing Price Appreciation

The index of housing price appreciation is a measure of relative change in property values over time. We construct housing price appreciation indices for four different property types (condominiums, one-family buildings, two- to four-family buildings, and multifamily rental buildings with five or more units) for New York City as a whole and for each borough and community district. Estimating price indices separately for different types of properties allows for different market valuations and fluctuations within each property type. However, since many community districts lack a sufficient number of properties of certain types (for example, there are very few single-family buildings in the Financial District) to appropriately estimate reliable housing price indices for those property types, we do not report a price index for all property types for each community district.

The data used to construct the price index come from two sources, both obtained from the New York City Department of Finance (DOF). The first data set is an annual sales file, which we receive under an exclusive arrangement. The second data set is the Automated City Register Information System (ACRIS) sales data, which is available online from the Department of Finance. Both data sets contain information on address, price, and date of sale for all transactions involving sales of apartment buildings, condominiums, and single- and multifamily homes in New York City between 1974 and 2014. While the ACRIS data are updated daily, the system contains less information on the circumstances of the sale than the annual sales file. The ACRIS data are used only if the sale is not recorded by the time we receive our annual sales file. The repeat sales price indices are created using statistical regression techniques. Economists use two basic approaches to estimate housing price indices: the hedonic regression (which tries to predict prices based on measurements of the quality of the unit as well as conditions of the surrounding neighborhood) and the repeat sales method. Both of these approaches estimate temporal price movement controlling for the variation in the types of homes sold from period to period. Each method has its own strengths and weaknesses.

The repeat sales method controls for housing characteristics by using data on properties that have sold more than once. An attractive feature of this method is that, unlike the hedonic approach, it does not require the (necessarily imperfect) measurement of housing unit quality; it only requires that the quality of individual units in the sample did not vary over time. The most important drawback of the repeat sales method is that it is based only on properties that have sold more than once in the study period. Moreover, properties that have been sold more than once may not be representative of all properties in the market, raising concerns about sample selection bias. However, as the index period lengthens, the proportion of properties that have changed hands multiple times increases. This reduces sample selection bias but exacerbates another problem: Case and Shiller (1989) present evidence that homes with longer intervals between sales have more volatile changes in price, since the longer the time between sales, the more likely it is that some external shock to the property itself or the surrounding

2 In this year's analysis, we define middle-income as 120 percent of AMI. Note that Table 1 displays HUD income limits and maximum affordable rents at 130 percent of AMI.

buildings has, independent of the price level of housing in the neighborhood, significantly affected prices. This report overcomes most of the problems associated with the repeat sales method. Specifically, the data set used here is quite large, so we lose little precision by eliminating properties that sold only once: in the 40 years captured by our data, 61 percent of residential lots changed hands at least twice by the end of 2012. In addition, we use the three-step procedure suggested by Case and Shiller (1989) and modified by Quigley and Van Order (1995) to account for the possibility that price changes are more volatile (that is, have higher variances) for properties that are sold less frequently.

In the first stage, the difference between the log price of the second sale and the log price of the first sale is regressed on a set of dummy variables, one for each year in the sample except for the base year (2000, when our index is set to equal 100[3]). For each pair of sales for a property, the dummy variables have values of +1 for the year of the second sale, -1 for the year of the first sale, and zeros otherwise. In the second stage, we calculate the squared difference between the sale price predicted by the first stage and the actual sale price and regress it on a constant term, the time interval between sales, and that time interval squared. This allows us to predict the variance of the differences between the prices predicted by the stage-one regression and the actual prices. In other words, we can predict how reliably the change in prices for a single property reflects price changes for properties overall. In the third stage, the stage-one regression is re-estimated by generalized least squares, weighting each observation by the inverse of the square root of the variance predicted by the stage-two regression. Essentially, we give lower weight to price changes for properties that, because there was a large time interval between sales, are more likely to reflect some fundamental change in the quality of the property itself or the immediately surrounding area and thus less likely to accurately reflect changes in the housing market overall.

Mortgage Lending Indicators

The Home Mortgage Disclosure Act (HMDA) requires financial institutions with assets totaling at least $42 million as of 2013 to report information on loan applications and originations if they have originated or refinanced any first-lien home purchase loans on one- to four-family properties (including condominium and co-op units) in the previous year. Thus, the HMDA data capture most, but not all, one- to four-family residential mortgage lending activity. We use this dataset to calculate the home purchase loan rate, the refinance loan rate, and a number of derivative indicators. All figures in our analysis are based on non-business-related loans on owner-occupied, one- to four-family properties (including condominiums). We exclude from our analysis any loans for manufactured or multifamily rental housing (with five or more units), loans on properties that are not owner-occupied, and any loans deemed to be business related (classified as those loans for which a lender reports an applicant's ethnicity, race, and sex as "not applicable").

The loans that we consider constituted about 85 percent of all loan originations in New York City in 2013. Since 2004, HMDA has required lenders to report when the spread between the annual percentage rate (APR) of a loan and the rate of Treasury securities of comparable maturity is greater than three percentage points for first-lien loans and five percentage points for junior-lien loans. In this report, all loans with an APR above this threshold are referred to as *higher-cost loans*. Loan applicants were assigned to a racial/ethnic group for purposes of our research based on the first reported race of the primary applicant. However, if the applicant reported his or her ethnicity as "Hispanic" the applicant was classified as Hispanic, regardless of the applicant's reported race. When an applicant provided information to the lender via mail, Internet, or telephone and did not provide information on their race, we assigned those loans to the "not reported" racial category. These loans were included in our city- and borough-level analyses but were omitted when calculating racial shares in the New York City section in Part 3.

Notices of Foreclosure (*Lis Pendens*)

We receive data on *lis pendens* (LP) filings from a private vendor, Public Data Corporation. An LP may be filed for a host of reasons unrelated to a mortgage foreclosure, so we use a variety of screening techniques to identify only those LPs related to a mortgage. These techniques include searching for words within either of the party names and dropping any LPs that relate to a tax lien or a mechanic's lien, or that are originated by a government agency. If the same property receives any additional LPs within 90 days of the initial LP, the additional LPs are not included in our rate to avoid counting the same foreclosure twice.

3 Note that for Part 1 of this year's report, we set the base year to 1990.

Properties that Entered REO

The data for this indicator come from two sources—LPs from Public Data Corporation and residential sales data from the New York City Department of Finance (DOF). Each of these datasets identifies properties using a unique borough, block, and lot number (BBL). Starting with the set of all LPs, we use BBLs to match each LP issued since 1993 with the most recent sale of that property prior to the LP (if the sale happened in 1974 or later). We then match the LP to any sales that occurred within three years from the date of the LP, and assume that the first such sale was undertaken in response to the foreclosure filing. To identify transfers into REO, we search the grantee name field of the first sale after the LP for the word "bank" or the name of any large bank or subsidiary. Finally, we check if the name of the grantee matches the name of the LP servicer. If this is the case we classify the sale as a transfer into REO.

Units Authorized by New Residential Building Permits

This indicator measures the number of residential units in proposed developments approved by the New York City Department of Buildings (DOB). We compile this indicator from job filings and permit approvals from DOB, which are publicly available on DOB's website for full years starting in 2004.[4] In New York City, developers file a job with DOB early in the development process. These records include many details about development projects including its extent (for example, if a project is a new building or alters an existing one) and, for residential projects, the number of housing units it will contain when complete. Because developers can file jobs long before DOB allows construction to begin, and our source of job filings rarely includes the date that a project is fully permitted, we must also collect permit data. Permits, which are associated with jobs, represent partial or entire approvals of development projects. Permits allow us to count only the projects in which DOB has approved structural work, so construction of those buildings is likely to occur. Because permits lack certain information about projects—the number of proposed housing units, in particular—we must merge some detail from jobs to permits. We consider only permits that meet the following criteria:

- The project will result in a new building (job type is "NB");
- The permit authorizes structural work (permit type is "NB");
- The development includes residential uses;
- The permit does not renew a previously approved permit (filing status is "initial");
- No other permit was filed for the same site during the previous calendar year.

When multiple permits on the same site (with the same building identification number, or BIN) meet these criteria, we count just the most recently issued permit. Thus, each permit we retain should represent a unique residential building project. The matching process for permits and jobs is somewhat imperfect. We are able to link most but not all permits to their associated jobs, because our data source does not include all job filings. When we cannot find a permit's matching job, we instead match the permit to the most recently filed job on the same BIN as the permit, as long as the job was filed no more than four years before the permit, and the job includes the number of units proposed for the site. In 2015, we counted 1,887 permits approved for new residential buildings; of that number, we matched 1,658 permits to their associated job and 172 permits to a recently filed job on the same site. We could not match 57 permits to jobs and therefore did not find the number of units proposed for those developments. Accordingly, our measure may somewhat understate the number of units in the construction pipeline.

Housing Unit Weighting Formula

Several indicators included in this report are provided at geographic levels other than the community district level—such as police precincts, school districts, or zip codes. We aggregate data to the community district level, weighting observations by the distribution of housing units. For instance, the New York City Police Department (NYPD) reports crime data at the precinct level, but we report crime rates at the community district level. When aggregating crime rates from the 76 police precincts to the 59 community districts, we first calculate the crime rate for each precinct. If a community district is entirely within one precinct, then that rate is directly used for the community district. If a community district contains portions of more than one

4 Available at http://www.nyc.gov/html/dob/html/codes_and_reference_materials/foilmonthly.shtml.

precinct, we weight the crime rate from each precinct based on the share of all housing units within the community district that are in that particular precinct. For example, if community district 1 contains three precincts, A, B, and C, and of the 100 housing units in community district 1, 50 are in precinct A, 30 are in precinct B, and 20 are in precinct C, then the crime rate from precinct A would have weight 50/100, the rate from precinct B would have weight 30/100, and the rate from precinct C would have weight 20/100. The crime rate for community district 1 would thus be given by:

$$rate_{CD1} = rate_A * .5 + rate_B * .3 + rate_C * .2$$

Halfway through 2013, a new precinct (the 121st) was created in Staten Island from portions of the 120th and 122nd precincts. Before weighting crime rates from the precinct to the community district level for 2013, we estimated the number of crimes for the 121st district as if it had existed for all of 2013 by multiplying the NYPD data for that precinct (which represented six months of data) by two. We then subtracted that number of crimes from the 120th and 122nd precincts based upon the ratio of crimes attributed to each precinct from the first half of 2013.

Calculating Distances to Parks

For New York City, each borough, and each community district, we report the percentage of housing units within one-quarter mile of a park. To calculate this figure, we first obtained a shapefile from the New York City Department of Parks and Recreation (DPR) (updated in September 2014 and available online through New York City's open data portal) describing the geographies of "functional parkland" overseen by the department. We then combine this with a shapefile we received from the New York State Office of Parks, Recreation, and Historic Preservation containing the geographies of state-owned parks. Any park the city categorizes as "undeveloped," a "lot," a "mall," a "parkway," or a "strip" is excluded from the analysis, as are parks smaller than a quarter of an acre. Because neither the city's nor the state's datasets contain information on the location of park entrances, we identify entrance points along each park's perimeter that constitute our best approximation of actual park entrances and then calculate walking distances from those entrance points. For parks with an area of less than two acres, we assume each vertex of the park polygon approximates a park entrance; since these parks are small, the actual location of entrances does not have a large effect on the walkshed (that is, the area reachable by walking a quarter mile or less along pedestrian rights-of-way starting at any of a park's entrance points). For parks of two acres or larger, the vertices may be too far apart to realistically approximate actual park entrances; for example, the four corners of Central Park are a very poor estimation of the entrances to the park. Thus, we instead find all the intersections of pedestrian rights-of-way that fall within 150 feet of the perimeter of these larger parks to approximate the entrance points. We obtained the pedestrian rights-of-way data from the New York City Department of City Planning's LION geodatabase of public streets. After we generate approximate park entrance points, we use Esri ArcMap's Network Analyst tool to generate walksheds estimating the areas along pedestrian rights-of-way that are located within a quarter mile of a park entrance point. In ArcMap we then select all building lots (which we get from the New York City Department of City Planning's MapPLUTO data) that fall within these walksheds and sum the total number of residential units on such lots and divide that number by the total number of residential units in a given geographic area.

Calculating Distances to Subways

For New York City, each borough, and each community district, we report the percentage of housing units within one-half mile of a subway station or rail entrance. To determine walking distances, the NYU Furman Center uses the New York City Department of City Planning's LION geodatabase of public streets to create network buffers of streets with pedestrian rights of-way within one-half mile of a subway entrance. Using geographic information systems (GIS), we then selected the lots that fell within this network buffer. We used a dataset of station entrances in the Bronx, Brooklyn, Manhattan, and Queens from the Metropolitan Transit Authority (MTA) through NYC DataMine. This dataset includes the following Metropolitan Transit Authority (MTA) constituent agencies: New York City Subway, Long Island Rail Road, and Metro-North Railroad. For the Staten Island Railway, we estimated station entrance locations using a variety of GIS techniques including current satellite imagery. Amtrak, PATH and New Jersey Transit stations are implicitly included in this calculation because their stations are co-located with stations within the systems named above.

Aggregating Student Performance

In a deviation from previous years' editions of the *State of New York City's Housing and Neighborhoods*, when reporting student performance at the city, borough, and community-district level, we use school-level proficiency rates; in prior years, we used a housing unit weighting formula to weight school-district level performance indicators to the community-district level.

The New York State Education Department publishes school-level proficiency rates every year. We joined the proficiency data with a school facilities shapefile provided by the New York City Department of City Planning's Bytes of the Big Apple website, which also includes the community district the school falls into. We removed private and charter schools and then summed up the number of fourth graders scoring "proficient" in math and English language arts in schools in a given community district, and the number of fourth grade students who were tested in each subject in that community district. We use those aggregates to calculate proficiency rates at the community district level. Since students can attend schools outside of their community district (for example, if their school zone extends beyond the borders of their community district), the student performance indicators provide information about the performance of students who attend schools in that neighborhood, rather than the performance of students who *live* in that neighborhood.

Metropolitan Transportation Authority Indicators

Subway Ridership and Performance

The Metropolitan Transportation Authority (MTA) published total annual ridership data for each station for the years 2009–2014. The MTA does not produce shapefiles with subway stops (although they do produce a file with each individual entrance and exit), but they release route and station data in the GTFS format. Steven Romalewski at the Center for Urban research has converted the GTFS data into a shapefile,[5] which we manually match to the performance data.

The MTA has published a list of stations that have been at least partially closed (that is, closed for a substantial period, rather than just during nights or weekends) between 2009 and 2014.[6] We remove from our analysis any station that was closed in one direction for more than 60 days or in both directions for more than 30 days during either 2009 or 2014; 21 out of the 419 subway stations met those criteria. For stations that were closed for a small portion of either 2009 or 2014, we determine what the ridership likely would have been had the station not been closed by calculating the average ridership per day for the portion of the year that the station was open and multiplying that rate by 365 days.

Note that several stations underwent substantial rehabilitations during the years between 2009 and 2014, and such stations are included in our analysis as long as the closure did not extend into 2009 or 2014.

Subway Performance Data

The MTA releases performance data for each subway line. We focus on two specific indicators: on-time performance (OTP), and subway wait assessment (SWA). In order to visualize performance geographically, we associate each subway station with the lines serving it, and average the indicators for each line at that station. Thus, the value at a particular station represents the average performance of lines serving that station. For more information, see "MTA Subway Performance" in the Indicators and Definitions section.

Inflation Adjustments

Unless stated otherwise, when reporting dollar-denominated indicators, we adjust amounts to 2015 dollars using the Consumer Price Index for All Urban Consumers (Current Series) without seasonal adjustments from the US Bureau of Labor Statistics over all major expenditure classes for the New York City metropolitan area. This allows for more consistent comparisons across years for individual indicators.

5 Available at https://wfs.gc.cuny.edu/SRomalewski/MTA_GISdata/June2010_update/nyctsubwaystops_100627_curcleaned.zip

6 A full list of closures can be found here: MTA New York City Transit Temporary Station Closures (2009-2014), http://web.mta.info/nyct/facts/ridership/ridership_sub_statClosure.htm.

Index of Community Districts

The Bronx

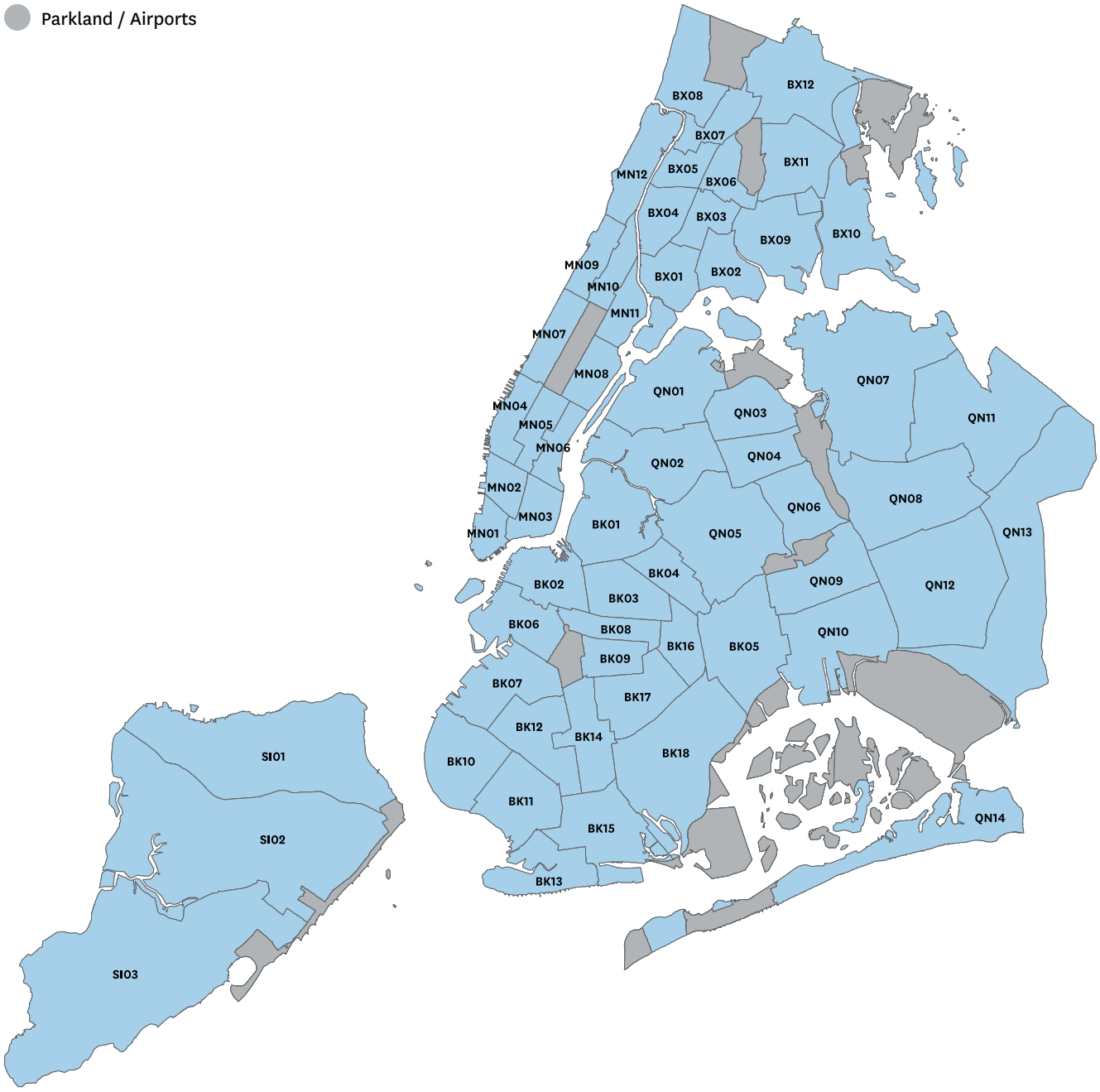
Brooklyn

Manhattan

Queens

Staten Island

New York City Community Districts



New York City Sub-Borough Areas