Methods

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 U.S. 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 beginning on page 142.

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. Half of the community board's members are appointed by the borough president; half are nominated by the City Council members who represent the 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, the Furman Center designates 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 Community Board 2 in Brooklyn.

SUB-BOROUGH AREA (SBA)

Sub-borough areas are geographic units created by the U.S. Census Bureau for the administration of the New York City Housing and Vacancy Survey and were designed to have similar boundaries to those of the community districts. These same areas are also defined by the U.S. Census Bureau as Public Use Microdata Areas (PUMAs) so we are able to use the two terms interchangeably.

Because sub-borough areas are constructed from census tracts, their boundaries do not coincide precisely with community district boundaries. However, they are similar enough that we use them interchangeably throughout this report. There are 59 community districts in New York City but only 55 sub-borough areas. The U.S. 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, Morrisania/Crotona (BX 03) and Belmont/East Tremont (BX 06) in the Bronx, the Financial District (MN 01) and Greenwich Village/Soho (MN 02) in Manhattan, and Clinton/Chelsea (MN 04) and Midtown (MN 05) in Manhattan.

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 the measure is for a quality that one might think is "best" if lower. When possible, we rank all 59 community districts, however, because data for several indicators—including all indicators drawn from U.S. Census Bureau sources—are only available at the sub-borough area level. We can only rank the 55 sub-borough areas with respect to these indicators. In addition, a few indicators are not available for all neighborhoods so we provide rankings for a subset of neighborhoods. For instance, the Furman Center only reports the index of housing price appreciation at the community district level for the predominant housing type in that district. Therefore, the rankings for these indicators come from a substantially reduced subset of the community districts.

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 U.S. Census Bureau. These sources are described below along with a discussion of issues of comparability across sources.

DECENNIAL CENSUS (CENSUS)

From 1970 to 2000, the decennial census consisted of two parts: the "short form" that collects information from every person and about every housing unit in the country, and the "long form" of additional questions asked 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 the City*, we use data from the decennial census short and long forms to derive demographic, economic, and housing measures for the year 2000. To create most of these indicators, we use summary census data reported at the city, borough, and sub-borough area levels.

In March of 2011, the Census Bureau released numbers from the 2010 decennial census for the five boroughs and for the city as a whole. We use these data to calculate the population, population density, housing units, racial/ethnic share, and racial diversity index at the city and borough levels. Whenever we report data from the 2010 decennial census, we do not report 2008 or 2009 American Community Survey data.

AMERICAN COMMUNITY SURVEY (ACS)

The American Community Survey is a relatively new 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 Census "long form" covered one out of every six housing unit addresses while the ACS only covers one in 40 housing units each year. The U.S. 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 U.S. Census Bureau began releasing three-year rolling estimates for all geographic areas with populations of 20,000 or more. In December 2010, the U.S. Census Bureau began releasing five-year rolling estimates for all geographic areas including census tracts. In this edition of the *State of the City*, we use ACS data to generate the same statistics we obtained from the 2000 decennial census, but for the years 2008 and 2009. Most of the indicators in this edition are derived from summary-level data reported by the U.S. Census Bureau for PUMAs which, as discussed above, are identical to New York City's sub-borough areas. Summary-level data are also reported at the borough and city levels. Because each PUMA in New York City has at least 100,000 residents, reliable annual estimates are available for each PUMA from the ACS. In this edition of *State of the City* we use annual estimates for almost all of the data we get from the ACS. One exception is the rental vacancy rate, for which we use the three-year estimate (see the section below for more details). We also use the three-year estimate to describe the racial composition in the following sub-borough areas: Highbridge/South Concourse (103) and Sheepshead Bay/Gravesend (215) because 2009 data were not available for those areas.

PUBLIC USE MICRODATA SAMPLE (PUMS)

While most indicators that draw on U.S. Census Bureau data are calculated using values that are already available at a given geography, the Furman Center calculates some indicators by aggregating household-level data to the required geography. The U.S. Census Bureau makes household-level data available in Public Use Microdata Samples, which are censored extracts from the confidential microdata that the U.S. Census Bureau uses in its own calculations.

The Furman Center uses PUMS data to calculate the income diversity ratio and several indicators in the State of New Yorkers section.

NEW YORK CITY HOUSING AND VACANCY SURVEY (HVS)

The Housing and Vacancy Survey is conducted every three years by the U.S. Census Bureau under contract with the City of New York. The New York City Department of Housing Preservation and Development spon-

sors and supervises the HVS. The primary purpose of the HVS is to satisfy the City's statutory requirement to measure the rental vacancy rate in order to determine if rent regulation will continue. In addition to the housing unit information, a limited set of data are also collected about the household and the individual answering the questionnaire.

In this edition of the *State of the City*, we use HVS data to construct one indicator that is specific to New York City and therefore not captured in the ACS: the percentage of rental units that are rent regulated.

COMPARISONS BETWEEN CENSUS BUREAU PRODUCTS

The U.S. 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 methodology might affect the comparability of indicators that we report in the *State of the City* over time. More information about comparability between U.S. Census Bureau data sources is available at: http://www.census.gov/acs/www/guidance_for_data_users/comparing_data/

Sampling

Because both the ACS and HVS are sample surveys, not censuses, all data derived from them are estimates, not exact counts. The ACS sample includes approximately three million housing units nationwide, including about 66,000 in New York City; the HVS samples 18,000 housing units. The sample for the HVS is designed primarily to achieve acceptable reliability in estimating the "vacant available for rent" rate for the entire city, so estimates for smaller geographic units such as sub-borough areas are subject to potentially large sampling errors. Readers should treat all estimates with some skepticism and be aware that the true value may differ significantly from the reported estimate. This is especially important when comparing small year-to-year changes in the ACS.

Income

Question construction and data collection for income information differs between the decennial census and the ACS. The 2000 census asked for the respondent's 1999 income; thus incomes reported in 2000 are all for one fixed period of time (calendar year 1999). The ACS, by contrast, asks for the respondent's income over the "past 12 months" and as this information is collected on an on-going monthly basis, these figures are not directly comparable. The U.S. 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 have reference months in common; thus comparisons of income data between adjacent ACS years (2008 and 2009) should not be interpreted as precise comparisons of economic conditions in those years.

Indicators affected by the income methodology issues are: income diversity ratio, median household income, poverty rate, and poverty rate by age.

Note that for comparison purposes, we adjust all dollar amounts reported in this report to 2010 dollars.

Rental Vacancy Rate

Nearly two thirds of the sub-borough areas in New York City lacked enough sample observations to calculate a rental vacancy rate for at least one year of ACS data. However, all had sufficient observations to calculate a three-year average of the rental vacancy rate. Thus, on the community district pages, for the rental vacancy rate only, we report a three-year average rental vacancy rate for 2007–2009. We still report annual rental vacancy rates on the borough and city pages, but the reported value for community districts cannot be directly compared to any one year of borough or city data.

INDICATOR NOTES

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, single-family homes, 2–4 family homes, and 5+ unit rental apartment buildings) 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. Due to insufficient data, we report the price indices only for the predominant property type at the community district level and at the two predominant property types for each borough.

The data used to construct the price index come from two sources, both obtained from the New York City Department of Finance. The first dataset is an annual sales file which we receive under an exclusive arrangement. The second dataset is the Automated City Register Information System (ACRIS) sales data which is available online from the Department of Finance. Both datasets contain information on address, price, and date of sale for all transactions involving sales of apartment buildings, condominium apartments and single- and multi-family homes in New York City between 1974 and 2010. While the ACRIS data are updated daily, the system does not contain data for sales in Staten Island. Therefore, the annual sales file is more complete. 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 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 methodology 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 measurement of house quality; it only requires the quality of individual houses in the sample to be time invariant. The most important drawback of the repeat sales method is that it fails to use the full information available in the data. In most datasets, only a small proportion of the housing stock is sold more than once; the data on single sales cannot be used. Moreover, properties that transact 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, more properties have changed hands more than once. This reduces sample selection bias but exacerbates a heteroskedasticity problem: Case and Shiller (1989) show evidence that price variability is positively related to the interval of time between sales because the longer the amount of time between sales, the more likely it is that the surrounding neighborhood has experienced an exogenous shock.

This report overcomes most of the problems associated with the repeat sales method. Specifically, the dataset used here is quite large, so we lose little precision by eliminating properties that sold only once. Moreover, because we have sales data over such a long period (37 years), more than 61 percent of residential lots have changed hands at least twice. Finally, we use the three-step procedure suggested by Case and Shiller[1] (1989) and modified by Quigley and Van Order[2] (1995) to account for the possibility of time-dependent error variances.

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 time period in the sample (a year, in this case) except for the first. 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, the squared residuals from the first stage are regressed on a constant term, the time interval between sales, and the time interval squared. The fitted value in the stage-two regression is a consistent estimate of the error variance in the stage-one

1 Case, K.E. and R.J. Shiller. 1989. "The Efficiency of the Market for Single Family Homes." American Economic Review, 79, p.125-37.

2 Quigley, J.M. and R. Van Order, 1995. "Explicit Tests of Contingent Claims Models of Mortgage Default." Journal of Real Estate Finance and Economics, 11(2), p.99-117.

regression. In the third stage, the stage-one regression is re-estimated by generalized least squares, using the inverses of the square root of the fitted values from the stage-two regression as weights.

MORTGAGE LENDING INDICATORS

The Federal Home Mortgage Disclosure Act (HMDA) requires financial institutions with assets totaling $39 million or more to report information on loan applications and originations if they have originated or refinanced any home purchase loans on a 1–4 family properties in the previous year. Thus, the HMDA data capture most, but not all, 1–4 family residential mortgage lending activity. The Furman Center uses 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 1–4 family, non-business-related loans. We exclude from our analysis, except when expressly noted, any loans for manufactured or multi-family housing (5+ families), 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 more than 80 percent of all loan applications in New York City in 2009.

Beginning in 2004, HMDA requires 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 high 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 for our *State of New Yorkers* section.

For a detailed look using HMDA data at national lending in 2009, see the Furman Center's report at http://furmancenter.org/files/HMDA_2009_databrief_FINAL.pdf.

NOTICES OF FORECLOSURE

The Furman Center collects 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 the Furman Center uses 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, a mechanic's lien, or 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.

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. 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.

POPULATION WEIGHTING FORMULA

Several indicators included in this report are provided at geographic levels other than the community district level such as school districts or zip codes.We aggregate data to the community district level, weighing observations by the distribution of housing units.

For instance, when aggregating the student proficiency rates from the 32 school districts to the 59 community districts, we first calculate the rate for each of the 32 school districts. If a community district only contains one school district then that rate is directly used for the community district. If multiple community districts fall within the same school district, we assign the same proficiency rate to each. If a community district contains more than one school districts, we weight each school district based on the number of housing units within the community district that are in that school district.

For example, if community district 1 contains three school districts A, B, and C, and of the 100 housing units in community district 1, 50 are in school district A, 30 are in school district B, and 20 are in school district C, then school district A would have weight 50/100, school district B would have weight 30/100, and school district C would have weight 20/100. The rate for community district 1 would be given by:

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

Since school district and community district boundaries are not coterminus, it is possible that the same school would be included in the calculation of two or more community districts. However, it would be weighted accordingly each time.

CALCULATING DISTANCE TO AMENITIES IN GIS

This report presents several indicators that show the percentage of housing units within a given walking distance to amenities.

To determine walking distances to amenities, the Furman Center used the NYC Department of City Planning's LION shapefile to create network buffers of streets with pedestrian rights-of-way within one half mile from a subway entrance and one quarter mile from the perimeters of parks. Using GIS, we then selected the lots that fall within this network buffer.

Subway/Rail Entrances

We use a database of station entrances in the Bronx, Brooklyn, Manhattan, and Queens from the Metropolitan Transit Authority through NYC DataMine. This dataset includes the New York City Subway system, Long Island Rail Road, and Metro-North Railroad. For the Staten Island Railway, we interpolate station entrances using a variety of GIS techniques including current satellite imagery. There are no Amtrak stations that are not colocated with other transportation services.

Parks

We access a database of all parks, playgrounds and greenstreets that are administered by the Department of Parks and Recreation through NYC Data Mine. Because our data on parks do not contain information on their entrances, we calculate walking distances from the nearest point along their perimeter. For parks with an area of 2.5 acres or less, we complete the analysis using only points at the corners of the parks perimeter. For parks larger than 2.5 acres, this would result in perimeter points that are too far apart. Instead, we use the intersections of pedestrian rights-of-way within 150 feet to approximate their perimeters. Parks that are less than one quarter of an acre are not considered.

INFLATION ADJUSTMENTS

When reporting dollar-denominated indicators, we adjust amounts to 2010 dollars using the Consumer Price Index for All Urban Consumers (Current Series) from the Bureau of Labor Statistics for all major expenditure classes for the NY-NJ-PA Metropolitan Statistical Area. This allows for more consistent comparisons across years for individual indicators. The inflation-adjusted values include median monthly contract rent, median household income, and median price per unit.