

NYC Geodatabase (nyc_gdb) Data Dictionary

Version: jan2019

Frank Donnelly, Geospatial Data Librarian, Baruch College CUNY


January 4, 2019

Abstract

This document provides metadata for the individual layers and tables in the NYC Geodatabase (nyc_gdb), a resource for mapping and analyzing city-level features and data in GIS. The database comes in two formats: a Spatialite geodatabase built on SQLite that can be used in open source software like QGIS and the Spatialite GUI, and a personal geodatabase built on MS Access that can be used in ArcGIS. The features and tables in each database format are identical, and this document describes them both using Dublin Core elements and terms. For a narrative description and tutorial on how to use the database, see the document "Introduction to the NYC Geodatabase (nyc_gdb)". The databases and associated documentation are available at <https://www.baruch.cuny.edu/confluence/display/geoportal/NYC+Geodatabase>.

Rights

Disclaimer: Every effort was made to insure that the data, which was compiled from public sources, was processed accurately for inclusion in the NYC Geodatabase. The creator, Baruch College, and CUNY disclaim any liability for errors, inaccuracies, or omissions that may be contained therein or for any damages that may arise from the foregoing. Users should independently verify the accuracy of the data for their purposes

The database and associated documentation are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike license CC BY-NC-SA <http://creativecommons.org/licenses/by-nc-sa/4.0/> . You are free to share and to adapt the work as long as you cite the source, do not use it for commercial purposes, and release adaptations under the same license.

Updates

The database will be updated biannually, with updates to individual layers or tables occurring in the summer (between July and August) or winter (between January and February), every ten years following the census (decennially), or irregularly if the data changes infrequently.

This twelfth version of the geodatabase, released in January 2019, supersedes the version released in January 2018. The following changes were made:

1. All of the ACS data tables for the 2012-2016 series were dropped, and were replaced with the 2013-2017 series for PUMAs, ZCTAs, and census tracts.
2. The latest data for business establishments and employment from the US Census ZIP Code Business Patterns was added for 2016: `b_zctas_2016biz_emp` and `b_zctas_2016biz_ind`. The source data was aggregated from the ZIP Code to the ZCTA level using a crosswalk table, `b_zips_to_zctas`. Data from 2015 was dropped from the database, and in the archives a short SQL script was created to update some mistakes in the 2015 industry table for ZCTAs 11694 and 11697 in Queens.

3. The `a_subway_complexes` layer was updated. Ridership data from 2007 to 2011 was carried over and left unchanged, data from 2012 to 2016 was replaced, and data for 2017 was added. The `a_subway_complexes_notes` table was updated with information to reflect station closures from 2007 to 2017. The MTA has not published official lists of closures since 2014, so subsequent information was gathered from MTA press releases and various sources on the internet. Complexes for three new stations on the 2nd Ave subway line in Manhattan were added, and missing ridership data from 2016 for the 138 St-Grand Concourse station in the Bronx was included in this release. The related `a_subway_stations` layer was updated to: add complex IDs for the new 2nd Ave stations, replace the old South Ferry Loop station with repaired South Ferry station on the 1 line, and mark the Cortlandt Street station on the 1 line as open.
4. The NYC Path train stations layer `a_path_stations` was updated with new ridership data for 2017.
5. All of the point features for city facilities (colleges, public libraries, hospitals, private schools, and public / charter schools) from May 2017 were dropped and replaced with data from November 2018. The primary key column for each table has been modified from previous versions and is now different for every table. The capacity field for the schools was dropped as the data has become ambiguous, and the utilization field for the schools was renamed enrollment. Compared to earlier iterations of the NYC GDB there is a larger number of public and charter schools as there may be duplicates between NYC and NYS sources. In addition, records for Staten Island public schools were missing valid identifiers and enrollment data, and thus were scrutinized to insure that valid school records were included. Data quality for the public and charter schools has become highly suspect and should be interpreted with caution.

The Database

The Spatialite database was built using version 4.1.1. The MS Access geodatabase was built using ArcGIS 10.1.

dc:title		NYC Geodatabase
dc:title		nyc_gdb
dc:creator		Frank Donnelly
dc:creator		Geospatial Data Librarian

dc:description	<p>The NYC Geodatabase (nyc_gdb) is a resource designed for basic geographic analysis and thematic mapping within the five boroughs of New York City. It contains geographic features and data compiled from several public sources. All of the features were transformed to share a common coordinate reference system (CRS) that is appropriate for the area: NAD 83 NY State Plane Long Island (feet); EPSG code 2263. Subsets of large features like water, greenspace, and public facilities were created and Census geographies like tracts, ZCTAs, and PUMAs were geoprocessed to create land-based boundaries. Census data from the 2010 Census, American Community Survey (ACS), and ZIP Code Business Patterns are stored in tables that can be easily related to geographic features. Transit and public facility point data were gathered from several city agencies and transformed into spatial data that can be used for reference or analysis for measuring distance, drawing buffers, or counting features within areas.</p> <p>The database contains many foundational map layers and data that can be readily used, but was also constructed so users could build on that foundation and extend it for their own projects. All of the boundaries are based on the 2010 Census, which allows users to easily add additional layers from the census TIGER files or to extend the study area beyond the five boroughs. The database also serves as an educational tool for introducing spatial databases and SQL.</p> <p>The dataset is appropriate for thematic and reference mapping at a city and borough level and for thematic mapping at a sub-borough level. While it can be used for creating detailed reference maps at a sub-borough level, it is not the ideal choice for this purpose, given the degree of generalization in the TIGER Line files (in terms of the detail of the line work for the coast line and the number of water and landmark features selected for inclusion). Users will have to judge for themselves based on their intended purpose.</p>	
dc:subject	xsi:type=dcterms:LCSH	Geographic Information Systems
dc:subject	xsi:type=dcterms:LCSH	Geodatabases
dc:subject	xsi:type=dcterms:LCSH	New York City-Geospatial data
dc:subject	xsi:type=dcterms:LCSH	United States-Census
dc:publisher		Baruch College, CUNY
dcterms:issued	xsi:type=dcterms:W3CDTF	2019-01
dc:type	xsi:type=dcterms:DCMIType	dataset
dc:format		sqlite
dc:format		msaccess
dc:identifier		https://www.baruch.cuny.edu/confluence/display/geoportal/NYC+Geodatabase
dc:identifier		nyc_gdb_jan2019.sqlite
dc:identifier		nyc_gdb_jan2019.mdb
dc:source		US Census TIGER / Line Shapefiles
dc:source		2010 Census Centers of Population
dc:source		NYC Dept City Planning Facilities Database (FacDB)
dc:source		MTA NYC Transit
dc:source		Port Authority NY NJ
dc:source		2010 Census
dc:source		2013-2017 American Community Survey
dc:source		2016 ZIP Code Business Patterns
dc:language	xsi:type=dcterms:ISO639-3	eng

dcterms:spatial	xsi:type=dcterms:Box	name=New York City; northlimit=272752.873521; east-limit=1067309.746197; southlimit=120053.504637; west-limit=913090.689878; units=ft; projection=NAD83 / New York Long Island (ftUS)
dcterms:license dc:rights		Attribution-NonCommercial-ShareAlike CC BY-NC-SA http://creativecommons.org/licenses/by-nc-sa/4.0/
dcterms:accrualperiodicity dcterms:accrualpolicy		biannual active

xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance | xmlns:dc=http://purl.org/dc/elements/1.1/ | xmlns:dcterms=http://purl.org/dc/terms/

A Tables

Objects that begin with the prefix "a" are geographic vector features that represent points, lines, and areas. The Census publishes updates to the TIGER Files each year to correct errors and update legal boundaries, but the statistical boundaries do not change until a new decennial census is conducted. With the exception of PUMAs, the 2010 Census geography (from the 2012 TIGER files) is fixed in this dataset and will not be updated until 2020. The release of PUMA boundaries and their subsequent use in the ACS is typically delayed for a few years following each decennial census. Each "a" table has a bcode field that indicates which borough the feature is located in. The bcode is the US Census ANSI / FIPS code for the county.

- 36005 - Bronx County (Bronx)
- 36047 - Kings County (Brooklyn)
- 36061 - New York County (Manhattan)
- 36081 - Queens County (Queens)
- 36085 - Richmond County (Staten Island)

Title	a_boroughs
Subject	Boroughs
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER county file, re-projected to local state plane and modified by subtracting a subset of the Census TIGER water layer (a_water_coastal) from it to create land-based boundaries. The unique ID is bcode; the FIPS state-county code.

Title	a_colleges
Subject	Colleges and Universities
Date	November 2018
Type	Point
Source	City Planning Facilities Database, New York City Department of City Planning
Source Link	http://www1.nyc.gov/site/planning/data-maps/open-data.page
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Biannual (winter & summer)
Description	This layer is an extract of NYC's Facilities Database (FacDB) that has been converted to a spatial layer. The data was taken "as is" and was not verified for accuracy or omissions. It includes all facilities that had a valid (not null) identifier from the NYS Education Department and that were coded as Colleges or Universities under Facility Subgroup. The unique ID is idagency; a field created by NYSED.
Title	a_facilities
Subject	Large Public Facilities
Date	2010
Type	Polygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER landmarks file, re-projected to local state plane and modified by subtracting a subset of the Census TIGER water layer (a_water_coastal) from it to create land-based boundaries. The largest public features: JFK and LaGuardia airports, the Brooklyn Navy Yard, Rikers, and the Psychiatric Center on Randalls Island, were selected to provide basic map reference and to cover up un-populated areas when making thematic maps. The feature for JFK airport was modified from the original file, which contained duplicate polygons. The mtfcc field contains Census MAF/TIGER Feature Class Codes that classify features by type. The unique ID is areaid; created by the Census.

Title	a_greenpace
Subject	Greenspace : Parks, Wildlife Areas, Cemeteries, Golf Courses
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER landmarks file, re-projected to local state plane and modified by subtracting a subset of the Census TIGER water layer (a_water_coastal) from it to create land-based boundaries. The largest and most culturally prominent features were selected to provide basic map reference and to cover up un-populated areas when making thematic maps. The NYC Dept of City Planning thematic maps for the 2010 Census were used for guidance in selecting features. Many of the features were modified from the original file: adjacent polygons were consolidated into larger ones, feature names were modified or updated, and in some cases polygons were created or reshaped. In particular, many of the islands in Jamaica Bay were not designated as greenspace in the 2010 Census landmarks files but were in the 2000 file. These features were copied from the 2000 file and inserted into this file. The mtfcc field contains Census MAF/TIGER Feature Class Codes that classify features by type. The unique ID is newid, created for this layer using the state county FIPS code and sequential numbers assigned to features in alphabetical order.
Title	a_hospitals
Subject	Hospitals
Date	November 2018
Type	Point
Source	City Planning Facilities Database, New York City Department of City Planning
Source Link	http://www1.nyc.gov/site/planning/data-maps/open-data.page
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Biannual (winter & summer)
Description	This layer is an extract of NYC's Facilities Database (FacDB) that has been converted to a spatial layer. The data was taken "as is" and was not verified for accuracy or omissions. It includes all facilities that had a valid (not null) identifier from the NYS Dept of Health and that were coded as Hospitals under Facility Type. The unique ID is idagency; a field created by NYS DOH.
Title	a_libraries
Subject	Public Libraries
Date	November 2018
Type	Point
Source	City Planning Facilities Database, New York City Department of City Planning
Source Link	http://www1.nyc.gov/site/planning/data-maps/open-data.page
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Biannual (winter & summer)
Description	This layer is an extract of NYC's Facilities Database (FacDB) that has been converted to a spatial layer. It includes most facilities that were coded as Public Libraries under Facility Type. Libraries that were listed as adult learning centers, reading and writing centers, or administrative buildings were removed from the dataset. With these exceptions, the data was taken "as is" and was not verified for accuracy or omissions. The unique ID is uid; a field created by the City.

Title	a_metro_counties
Subject	NYC Metropolitan Area Counties
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York Metropolitan Area
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER county file, re-projected to local state plane and modified by subtracting a subset of the Census TIGER water layer (a_water_coastal) from it to create land-based boundaries. The unique ID is countyid; the FIPS state-county code.
Title	a_path_stations
Subject	PATH Stations and Ridership
Date	2018
Type	Point
Source	NJ Transit
Source Link	https://njgin.state.nj.us/NJ_NJGINExplorer/index.jsp
Source	Port Authority of New York and New Jersey
Source Link	http://www.panynj.gov/path/statistics.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Annual (summer)
Description	This layer is a subset of the NJ Transit passenger railroad station shapefile provided through the NJ Geographic Information Network. This subset contains just the PATH train stations that are located in NYC, reprojected to NY state plane. Annual station ridership data was extracted from reports published by the Port Authority and appended to each station. Annual, average weekday, and average weekend ridership is provided for 2012 to 2017. Figures for average weekend ridership were calculated by taking the Saturday and Sunday totals and averaging them based on the number of weekend days that year. Neither the weekday nor weekend averages include holidays, which PANYNJ tabulates separately. The unique ID is atis_id, a field created by NJ Transit.

Title	a_pumas2010
Subject	Public Use Microdata Areas : PUMAs
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER PUMA file, re-projected to local state plane and modified by subtracting a subset of the Census TIGER water layer (a_water_coastal) from it to create land-based boundaries. PUMAs are statistical areas built from census tracts that are designed to have approximately 100k residents. PUMA boundaries from the 2010 Census were first used in the American Community Survey (ACS) 2012 series; earlier series used the 2000 Census boundaries. The Census began assigning official names (in addition to numbers) to PUMAs based on state and local government input beginning in 2010. The City decided to correlate the PUMAs with the names and numbers of community districts. The full, official Census name is stored in namelsad10, while the neighborhood name and community district numbers were broken out into separate fields called name and cd. This layer can be joined to "b" tables for mapping census data. The unique ID is geoid10; the FIPS state-puma code.
Title	a_roads
Subject	Roads
Date	2010
Type	Line
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a combination of the Census TIGER roads files for counties, re-projected to local state plane. Portions of roads in neighboring New Jersey were also included, so that bridges do not appear to halt in the middle of rivers (where state boundaries touch). The mtfcc field contains Census MAF/TIGER Feature Class Codes that classify features by type. A number of duplicate features in Brooklyn and Queens (errors in the line files) were deleted so that the id fields for the roads could serve as unique identifiers. The unique ID is linearid; created by the Census.

Title	a_schools_private
Subject	Private schools : elementary and secondary
Date	November 2018
Type	Point
Source	City Planning Facilities Database, New York City Department of City Planning
Source Link	http://www1.nyc.gov/site/planning/data-maps/open-data.page
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Biannual (winter & summer)
Description	This layer is an extract of NYC's Facilities Database (FacDB) that has been converted to a spatial layer. The data was taken "as is" and was not verified for accuracy or omissions. It includes all facilities that had a valid (not null) identifier from the NYS Dept of Education and that were coded as Non-Public K-12 Schools under Facility Subgroup, but were not coded as Satellite Sites under Facility Type. The enrollment field includes student enrollment for the fall of the 2017-18 academic year from the NYS Dept of Education. The unique ID is idagency; a field created by the NYSED.
Title	a_schools_public
Subject	Public schools : elementary and secondary
Date	November 2018
Type	Point
Source	City Planning Facilities Database, New York City Department of City Planning
Source Link	http://www1.nyc.gov/site/planning/data-maps/open-data.page
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Biannual (winter & summer)
Description	This layer is an extract of NYC's Facilities Database (FacDB) that has been converted to a spatial layer. With few exceptions the data was taken "as is" and was not verified for accuracy or omissions. It includes all facilities that were coded as Public K-12 Schools and Charter K-12 Schools under Facility Subgroup, and that had a valid agency id from either the NYCDOE or NYSED, except for Staten Island where most agency ID values were null. All records for Staten Island were included except for structures that were clearly not schools or that had facility types that were not present in the other four boroughs. Users should interpret this data with caution as there are likely duplicate records between the NYC and NYS sources and because the data from Staten Island is missing proper identifiers. The enrollment field includes student enrollment for the fall of the 2017-18 academic year from the NYC Dept of Education, but data for Staten Island was missing. Given the poor quality of this data no field could serve as a unique identifier, as the uid field has duplicates and the id agency field has nulls. A sequential integer number called rowid was created to serve as a key.

Title	a_subway_complexes
Subject	Subway Complexes and Ridership
Date	July 2018
Type	Point
Source	Developer Resources, MTA New York City Transit
Source Link	http://web.mta.info/developers/download.html
Source	Subway Ridership, MTA New York City Transit
Source Link	http://web.mta.info/nyct/facts/ridership/index.htm
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Annual (summer)
Description	This layer was originally created in August 2012 and has been subsequently updated. This layer is a subset of the a_subway_stations layer that has been combined with MTA statistics on ridership. Ridership data is not available for each individual subway station, as many stations are linked via common entrances and passageways where transfers are free, and because ridership data is not collected for the Staten Island Railway stations. a_subway_complexes was created by choosing an individual station from a_subway_stations to represent the entire complex, and modifying the station name and train fields appropriately. For a complex with two stations, the station with the most trains was selected. For a complex with three or more stations, the station that was in the geographic center was selected. a_subway_complexes should be used for mapping ridership data or for analysis that requires this data, and not for specifying actual station locations or measuring distances. There are 425 complexes, and the field station_ct indicates how many stations are part of a complex. Annual, average weekday, and average weekend ridership is provided for 2007 to 2017. The trains column only represents trains that served those complexes during the latest year for which ridership data is available; train service has changed over time and this is not reflected in this data layer. The unique ID is complex_id, which was created by alphabetizing the complexes by borough and station name and assigning a sequential number to a borough prefix.
Title	a_subway_complexes_notes
Subject	Subway Ridership Service Notes
Date	July 2018
Type	Table
Source	Temporary Station Closures (2007-2014), MTA New York City Transit
Source Link	http://web.mta.info/nyct/facts/ridership/ridership_sub_statClosure.htm
Source	Press Releases for NYC Transit, MTA
Source Link	http://www.mta.info/press-releases
Coverage	New York City
Accrual Periodicity	Annual (summer)
Description	This table contains notes on subway service disruptions by subway complex, to supplement the 2007 to 2017 ridership statistics in a_subway_complexes. complex_id is a foreign key that relates the notes to the subway complexes but it is not unique in this table, as a complex can have multiple disruption notes based on the direction of the disruption (i.e. uptown, downtown, or both). The table has been augmented with notes from additional sources. Since the MTA has not published closure notes with its ridership data since 2014, new information is gathered from MTA press releases and various media sources on the internet. The unique ID is id, which was a sequential number assigned to complex notes as they are added.

Title	a_subway_stations
Subject	Subway Stations
Date	November 2018
Type	Point
Source	Developer Resources, MTA New York City Transit
Source Link	http://web.mta.info/developers/download.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Annual (summer)
Description	This layer is a modified version of the MTA stops file that has been converted to a spatial layer and reprojected to local state plane. Four stops (D13, D20, N12, and R09) were dropped from the stops file because geographically they were duplicates of other stations. Two closed stations (Atlantic and Nassau on the SIR) and two non-service stations (South Ferry Loop and Broad Channel test track) were also dropped. The location of the Aqueduct Racetrack station on the A line was manually adjusted, as it was not located in the right location in the stops file. The 493 Stations represent distinct platforms served by specific trains. For example, 14th-St-Union Square is represented as three stations in distinct geographic locations based on train service: The L train, the N Q R trains, and the 4 5 6 trains. The complex_id field relates individual stations to shared complexes where platforms are connected and transfers are free in the a_subway_complexes layer. The multi field indicates whether the station is one of many stations in a complex. The names of the stations were taken directly from the MTA file, and may differ slightly from how they appear on official subway maps. The unique ID is stop_id, a field created by the MTA.
Title	a_tract_popcenters
Subject	Population Centers / Centroids by Census Tracts
Date	2010
Type	Point
Source	2010 Census Centers of Population by Census Tract
Source Link	http://www.census.gov/geo/reference/centersofpop.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is an extract of the 2010 Census Center of Population file that has been converted to a spatial layer and reprojected to local state plane. The population centers (aka population centroids) represent the center of a population's distribution within a census tract, which can be useful for operations like measuring the distance of a tract's population from some service or hazard. The notes field indicates whether a tract's population center falls outside the tract boundary in an adjacent tract or body of water; this can occur due to the population distribution and irregular geometric shape of some tracts. The pop2010 field contains the total population from the 2010 Census. This layer can be joined to "b" tables for mapping census data. The unique ID is tractid; the FIPS state-county-tract number code.

Title	a_tracts
Subject	Census tracts
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Source	2010 Geographies, New York City Department of City Planning
Source Link	http://www.nyc.gov/html/dcp/html/census/geo_notes.shtml
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER census tract file, re-projected to local state plane and modified by subtracting a subset of the Census TIGER water layer (a_water_coastal) from it to create land-based boundaries. Tracts are statistical areas built from census blockgroups that are designed to have an ideal size of 4,000 residents, with a range of 1,200 to 8,000. The nta fields indicate the Neighborhood Tabulation Area (NTA) to which the tract belongs; NTAs are statistical areas created by the NYC Dept of City Planning for presenting census data. This layer can be joined to "b" tables for mapping census data. Unique ID is tractid; the FIPS state-county-tract number code.
Title	a_train_stations
Subject	Train Stations : Metro North and Long Island Rail Road
Date	August 2012
Type	Point
Source	Developer Resources, MTA New York City Transit
Source Link	http://web.mta.info/developers/download.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Irregular
Description	This layer is a subset of the MTA stops file for Metro North and LIRR that has been combined and converted to a spatial layer and reprojected to local state plane. It contains just the stations that are in NYC. The trains field indicates whether the station is served by Metro North or LIRR. The unique ID is rail_id, a field created by attaching a prefix mn or lr to numbers assigned by the MTA.
Title	a_water_coastal
Subject	Coastal Water and Rivers
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York Metropolitan Area
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset and combination of the Census TIGER county water files for all counties in the metro region, re-projected to local state plane. Individual water features that were of significant collective size and that would be appropriate for city-level maps were selected from the master water feature set, with particular attention paid to areas around and along: the Atlantic Coast, Long Island Sound, Hudson River, East River, Harlem River, Arthur Kill, Hackensack River, Passaic River, New York Bay, Newark Bay, Raritan Bay, and Jamaica Bay. The individual water selections were exported out and unioned into one feature. a_water_coastal was used to modify the census statistical and landmark features in this dataset to create land-based boundaries.

Title	a_water_lakes
Subject	Water : Lakes and Reservoirs
Date	2010
Type	Polygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset and combination of the Census TIGER county water files for the five boroughs, re-projected to local state plane. Large, individual water features that represent lakes or reservoirs were selected from the master water feature set, to provide basic reference for thematic maps. The mtfcc field contains Census MAF/TIGER Feature Class Codes that classify features by type. The unique ID is hydroid, created by the Census.
Title	a_zctas
Subject	ZIP Code Tabulation Areas : ZCTAs
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER ZCTA file, re-projected to local state plane and modified by subtracting a subset of the Census TIGER water layer (a_water_coastal) from it to create land-based boundaries. ZCTAs are statistical areas designed to approximate USPS ZIP Codes. They are built from census blocks that are aggregated based on common postal addresses assigned to streets. There are some non-geographical ZIP Codes for which there are no corresponding ZCTAs. In NYC there are several small ZCTAs that cover non-residential office buildings or public facilities, primarily in Midtown Manhattan. ZCTAs may cross borough / county boundaries; the note field indicates the exceptions. In particular ZCTAs 11001, 11003, and 11040 are partially located in Queens, but a large majority of the land area and population are in Nassau County. This layer can be joined to "b" tables for mapping census data. The unique ID is zcta; the five-digit ZCTA / ZIP code.

B Tables

Objects that begin with the prefix "b" are tables that contain US Census data that can be related to the PUMA, tract, and ZCTA statistical features. The first part of the table name indicates the geography and the second part indicates the dataset and year. GEOID2 (the Census ANSI / FIPS code) is the unique ID in the ACS and 2010 Census tables, while ZCTA5 is the unique ID for the ZIP Business Pattern tables. Each dataset has a lookup table that relates the names of columns to the actual names of the variables / values.

Title	b_tracts_2010census
Title	b_tracts_2010census_pct
Title	b_zctas_2010census
Title	b_zctas_2010census_pct
Subject	2010 Census
Date	2010
Type	Table
Source	Summary File 1, Table DP01
Source Link	http://www.census.gov/2010census/data/
Coverage	New York City
Spatial	Census Tracts, ZCTAs
Accrual Periodicity	Decennial
Description	<p>These tables contain the demographic profile table (DP01) for each census tract and ZCTA, divided into two tables: one for total values and the other for percent totals (pct). The 2010 Census was a 100% count of the population taken on April 1st, 2010. The field names were created by the Census Bureau; the prefix indicates whether the value is a total (HD01) or percentage (HD02) and the suffix (S001) indicates the variable numbers. For totals (like total population, total housing units) the associated percentage is 100%. The field names can be tied to the actual names of the values in the index or look-up table, b_2010census_lookup. Some values in the index table have footnotes, indicated by a number in brackets beside the variable name. These footnotes are stored in the table b_2010census_footnotes. The ZCTA data includes ZCTAs 11001, 11003, and 11040, which are primarily located in Nassau County but are partially inside Queens. The data tables can be joined to "a" features for mapping census data. The unique ID is GEOID2, and the feature name is stored in GEOLABEL.</p>

Title	b_pumas_2017acs1
Title	b_pumas_2017acs2
Title	b_tracts_2017acs1
Title	b_tracts_2017acs2
Title	b_zctas_2017acs1
Title	b_zctas_2017acs2
Subject	American Community Survey
Date	2013-2017
Type	Table
Source	Tables DP02, DP03, DP04, and DP05
Source Link	https://www.census.gov/programs-surveys/acs/news/data-releases/2016/release.html
Coverage	New York City
Spatial	PUMAs, Census Tracts, ZCTAs
Accrual Periodicity	Annual (winter)
Description	<p>These tables represent subsets of the demographic profile tables for each PUMA, census tract and ZCTA, divided into two tables: one for social and economic data (acs1) and the other for housing and demographic data (acs2). The variables were selected to include the most common ones that would be of general interest while providing some degree of breadth, while not exceeding the functional column limits of the MS Access database (255). The American Community Survey (ACS) is a rolling sample survey of 3.5 million addresses published as 1 and 5-year estimates that are updated annually. This database includes just the 5-year estimates, as data for tracts and ZCTAs are only published for the 5-year series. Estimates are published at a 90% confidence interval with margins of error calculated for each value. The field names were created specifically for this dataset and group variables together by similar categories. Each variable has four parts identified with a suffix: "E" for the estimate, "M" for the margin of error (+/-) for that estimate, "PC" for the percentage of the total that the estimate represents, and "PM" for the margin of error for the estimated percentage. For totals (like total population, total housing units) the actual whole value is provided in the percentage field, and columns where percentages and their associated margins are not applicable (means, medians, rates) are left as NULL (following the conventions that the Census Bureau uses for these tables). The field names can be tied to the actual names of the values in the index or look-up table, b_2017acs_lookup. The ZCTA data includes ZCTAs 11001, 11003, and 11040, which are primarily located in Nassau County but are partially inside Queens. The data tables can be joined to "a" features for mapping census data. The unique ID is GEOID2, and the feature name is stored in GEOLABEL.</p>

Title	b_zctas_2016biz_emp
Title	b_zctas_2016biz_ind
Subject	Business establishments
Date	2016
Type	Table
Source	ZIP Code Business Patterns
Source Link	http://www.census.gov/programs-surveys/cbp.html
Coverage	New York City
Spatial	ZCTAs
Accrual Periodicity	Annual (summer)
Description	<p>These tables represent the ZIP Code Business Patterns (ZBP) data, divided into two tables: one for total establishments, employees, and payroll (emp) and the other for the number and percentage of business establishments by industrial classification code (ind). The Census Bureau compiles the ZBP data annually from the Business Register database, which contains a record for each known business establishment with paid employees in the US; an establishment is a single physical location at which business is conducted or services or industrial operations are performed. Payroll in the emp table is provided for the first quarter and entire year in \$1,000s of dollars. The field names in the ind table represent the two-digit North American Industrial Classification System (NAICS) codes; the sector names for each code are available in the b_zctas_2016biz_indcodes table. The records in these tables represent US Census ZCTAs and not USPS ZIP Codes; the original ZIP Code data from the ZBP was aggregated to the ZCTA level by summing the data for all ZIP Codes (included standard delivery areas, PO Box ZIP Codes, and large institutions) to the ZCTA where they are geographically located using the crosswalk table b_zips_to_zctas. The source table for relating ZIP Codes to ZCTAs was updated in 2015 to correct a small number of errors in assignments. The tables can be joined to "a" ZCTA features for mapping census data. In some instances records in the emp table may be missing values due to disclosure regulations and privacy rules; there are three note or flag columns that indicate how many establishments are missing from the tabulations. The data includes ZCTAs 11001, 11003, and 11040, which are primarily located in Nassau County but are partially inside Queens. The unique ID is ZCTA5.</p>

Title	b_zips_to_zcta
Subject	ZIP Codes and ZIP Code Tabulation Areas
Date	2015
Type	Table
Source	Missouri Census Data Center, Baruch CUNY
Source Link	http://mcdc.missouri.edu/allabout/zipcodes_2010supplement.shtml
Coverage	New York City
Spatial	ZIP Codes
Accrual Periodicity	Decennial
Description	This table is a crosswalk that relates USPS ZIP Codes with US Census 2010 ZCTAs. It was used to aggregate data from the Census ZIP Code Business Patterns from the ZIP Code to the ZCTA level and is provided in this database as a general reference for users. The US Census Bureau creates ZCTAs by aggregating census blocks based on the location of addresses in the blocks, to form geographic areas that share the same ZIP Code. As a result there are fewer ZCTAs than ZIP Codes, as ZIP Codes that lack a meaningful geographic area (typically includes clusters of PO Boxes or large organizations that process a lot of mail) are omitted. In this table all ZIP Codes are assigned to the ZCTA in which they are geographically located, to allow for aggregation. The data was downloaded from a master file from the Missouri Census Data Center and processed to create a subset for just New York City that includes a bcode. In Nov 2015 an updated version of this file was published to correct a number of assignment errors, and this file was used to recreate b_zips_to_zcta for the July 2017 release of the nyc_gdb, and to recalculate the data in the ZBP tables. ZCTA boundaries are updated every ten years following the decennial census. The data includes ZIPs & ZCTAs 11001 (includes ZIP Codes 11001 and 11002), 11003, and 11040, which are primarily located in Nassau County but are partially inside Queens. The unique ID is zipcode.

C Tables

Objects that begin with the prefix "c" are geographic vector features that represent areas. Other than having their coordinate systems reprojected, these files (features and attributes) have not been modified in any way from the original Census TIGER Line Shapefiles. They are included in the database for users who wish to depict actual, unmodified boundaries (that include both land and water) on maps for reference purposes. They are not intended for mapping data. The Census publishes updates to the TIGER Files each year to correct errors and update legal boundaries, but the statistical boundaries do not change until a new decennial census is conducted. With the exception of PUMAs, the 2010 Census geography (from the 2012 TIGER files) is fixed in this dataset and will not be updated until 2020. The release of PUMA boundaries and their subsequent use in the ACS is typically delayed for a few years following each decennial census.

Title	c_bndy_boroughs
Subject	Boroughs
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER county file, re-projected to local state plane. The unique ID is geoid; the FIPS state-county code.

Title	c_bndy_metro_counties
Subject	NYC Metropolitan Area Counties
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York Metropolitan Area
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER county file, re-projected to local state plane. The unique ID is geoid; the FIPS state-county code.
Title	c_bndy_pumas2010
Subject	Public Use Microdata Areas
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER PUMA file, re-projected to local state plane. PUMAs are statistical areas built from census tracts that are designed to have approximately 100k residents. PUMA boundaries from the 2010 Census were first used in the American Community Survey (ACS) 2012 series; earlier series used the 2000 Census boundaries. The unique ID is geoid10; the FIPS state-puma code.
Title	c_bndy_tracts
Subject	Census tracts
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER census tract file, re-projected to local state plane. Tracts are statistical areas built from census blockgroups that are designed to have an ideal size of 4,000 residents, with a range of 1,200 to 8,000. The unique ID is geoid; the FIPS state-county-tract number code.

X Tables

Objects that begin with the prefix "x" are geographic vector features that represent areas. These layers served as the source for several of the "a" layers in the database, which are subsets of the "x" layers. They are provided for users who wish to add additional features to the "a" layers.

Title	x_landmarks
Subject	Landmarks
Date	2010
Type	Polygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a combination of the county Census TIGER landmarks file for the five boroughs, re-projected to local state plane. Its features and attributes are unmodified from the original file. The mtfcc field contains Census MAF/TIGER Feature Class Codes that classify features by type. Features include parks, cemeteries, wildlife areas, airports, transit yards, and large public buildings like hospitals, museums, colleges, and city government offices. The a_facilities and a_greenpace layers were largely created using the landmarks geometry, with additions and modifications to features and attributes. This "x" layer is provided for users who wish to add additional features to the existing layers or to create new ones. The unique ID is areaid; created by the Census.
Title	x_nad83_boroughs
Subject	Boroughs
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York City
Spatial	NAD83, EPSG 4269
Accrual Periodicity	Decennial
Description	This layer is a subset of the Census TIGER county file, modified by subtracting a subset of the Census TIGER water layer (a_water_coastal) from it to create land-based boundaries. It is the only layer in the database that was not re-projected to local state plane; it uses the original geographic coordinate system that Census layers are projected in, NAD 83. It is provided for users as a frame of reference for plotting local longitude and latitude data, which can subsequently be saved as geographic features and converted to local state plane. The unique ID is bcode; the FIPS state-county code.

Title	x_water
Subject	Water
Date	2010
Type	Multipolygon
Source	US Census 2012 TIGER Line Shapefiles
Source Link	http://www.census.gov/geo/maps-data/data/tiger-line.html
Coverage	New York Metropolitan Area
Spatial	NAD83 New York Long Island (ft), EPSG 2263
Accrual Periodicity	Decennial
Description	This layer is a combination of the county Census TIGER water files for all counties in the metro region, re-projected to local state plane. Its features and attributes are unmodified from the original file with the exception of the include field - features with a value of 1 in this field were selected for inclusion in the a_water_coastal layer; that subset layer was unioned together into one feature and then used to modify the boundaries of all the areal "a" layers. The mtfcc field contains Census MAF/TIGER Feature Class Codes that classify features by type. This "x" layer is provided for users who wish to add additional features to the existing water layers or to create new ones. The unique ID is hydrdoid; created by the Census.

Other Tables

d_table : The object **d_ntas_2010census** is not a table, but a view that is included for the sake of example. It joins the a_tracts layer to the b_tracts_2010census table and groups basic population and housing data by Neighborhood Tabulation Areas (ntas) defined by the City.

z_table : The **z_metadata** table describes the name and source of all the tables in the database, along with the year the feature or table was last updated.

Other : Any remaining tables not mentioned in this documentation were automatically inserted when the databases were created. They form the core infrastructure of the geodatabase and should not be altered or removed unless the user is familiar with how they operate. The names of these tables differ between the SQLite and MS Access versions.