THE NYC GEODATABASE

An Open Source Spatialite Database

http://www.baruch.cuny.edu/geoportal/nyc_gdb/

Frank Donnelly, Geospatial Data Librarian

Baruch College CUNY

francis.donnelly@baruch.cuny.edu

Goals

- Provide general overview of Spatialite as a desktop spatial database for working in FOSS GIS

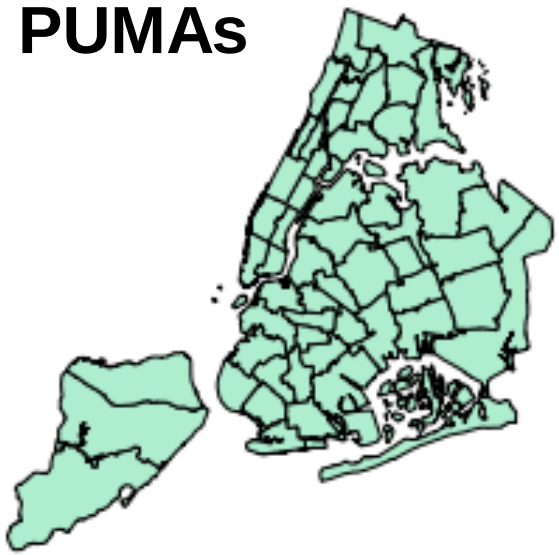
- Describe the NYC GDB in particular

NYC GDB

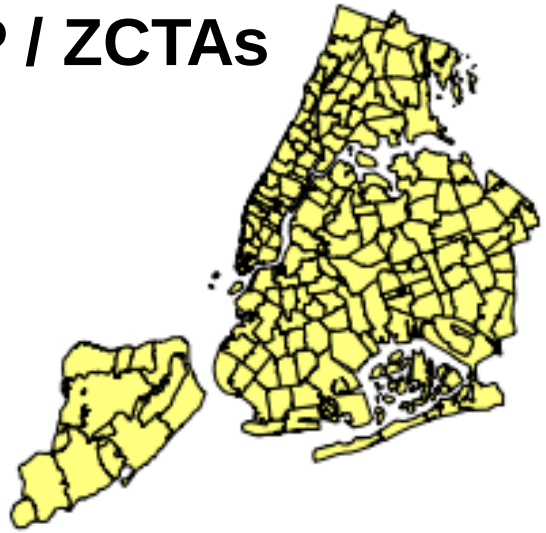
- Collection of city-level features and data tables for thematically mapping and analyzing neighborhood data, updated bi-annually, detailed tutorial and metadata
- Goal: provide new users with opportunity to learn about spatial databases, provide all users with basic and pre-processed features and data for city analyses
- Two formats: Spatialite and MS Access gdb for ArcGIS

Neighborhoods – 2010 Census TIGER

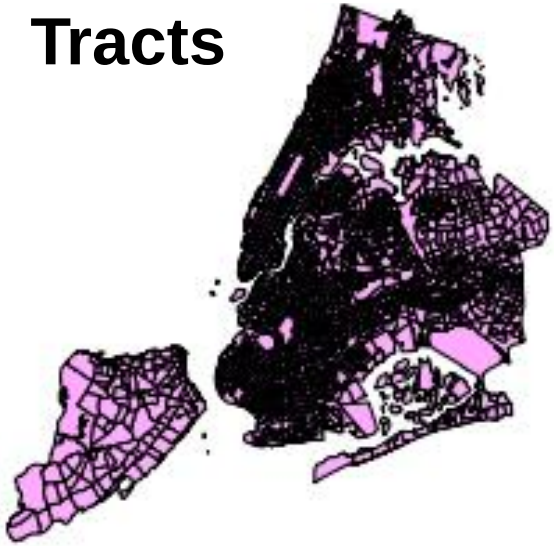
PUMAs



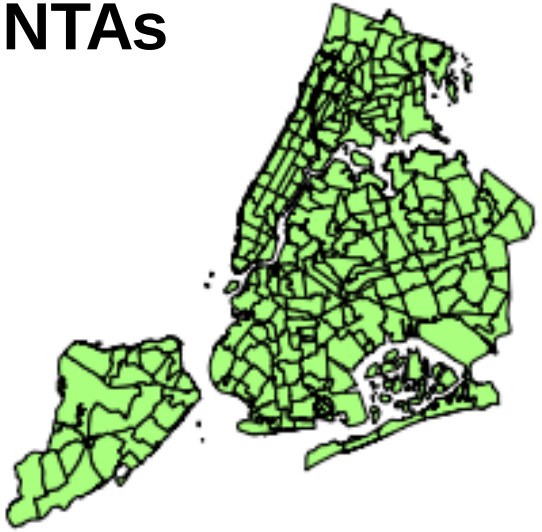
ZIP / ZCTAs



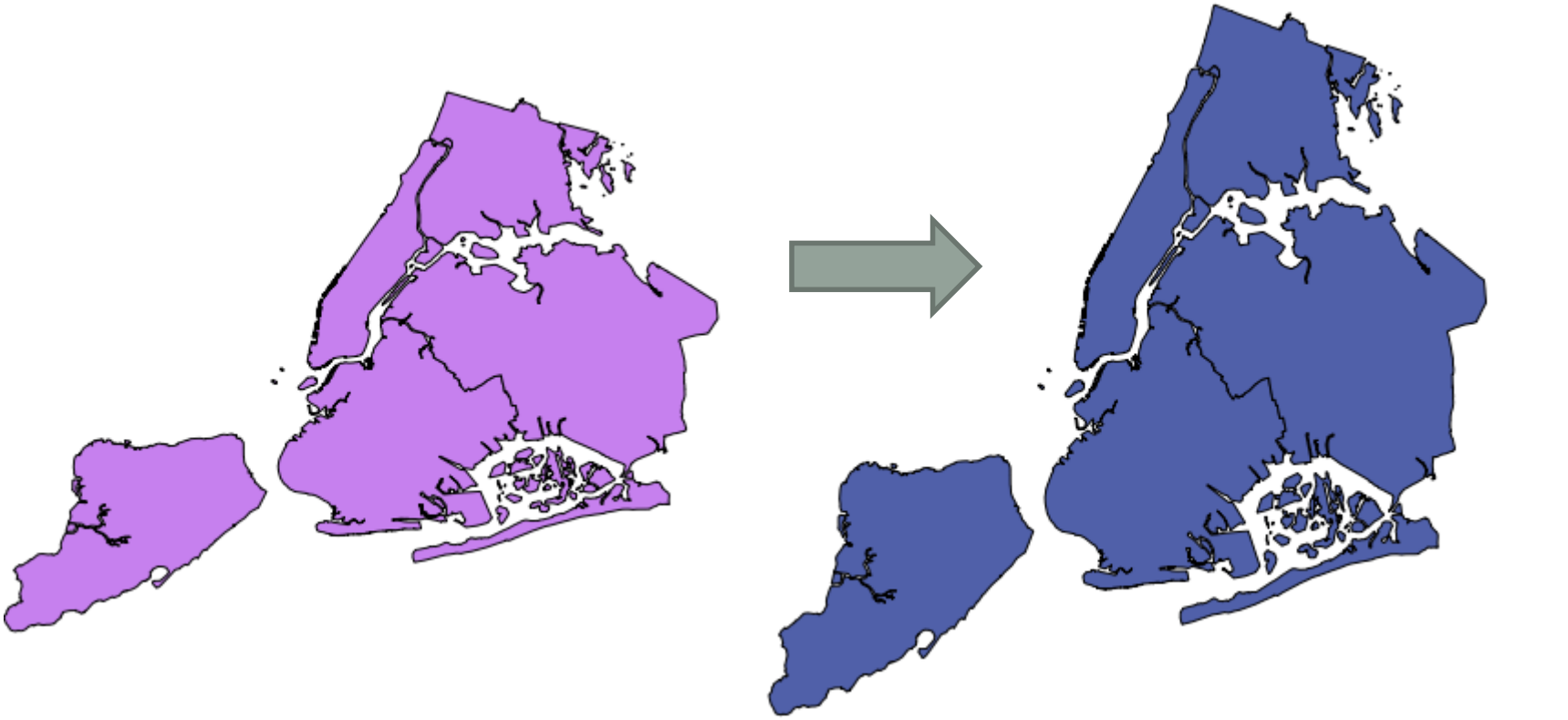
Tracts



NTAs

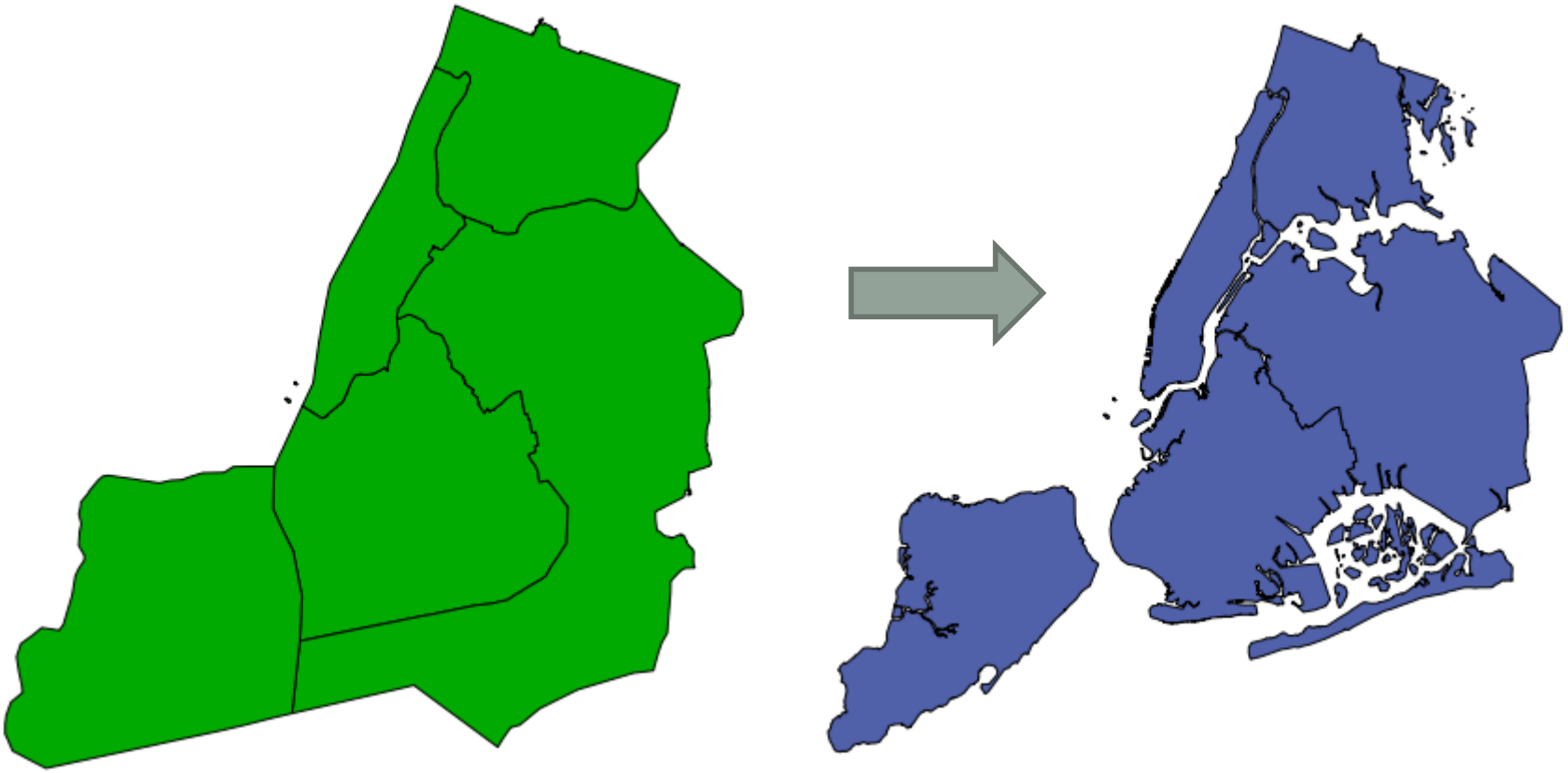


Re-projection

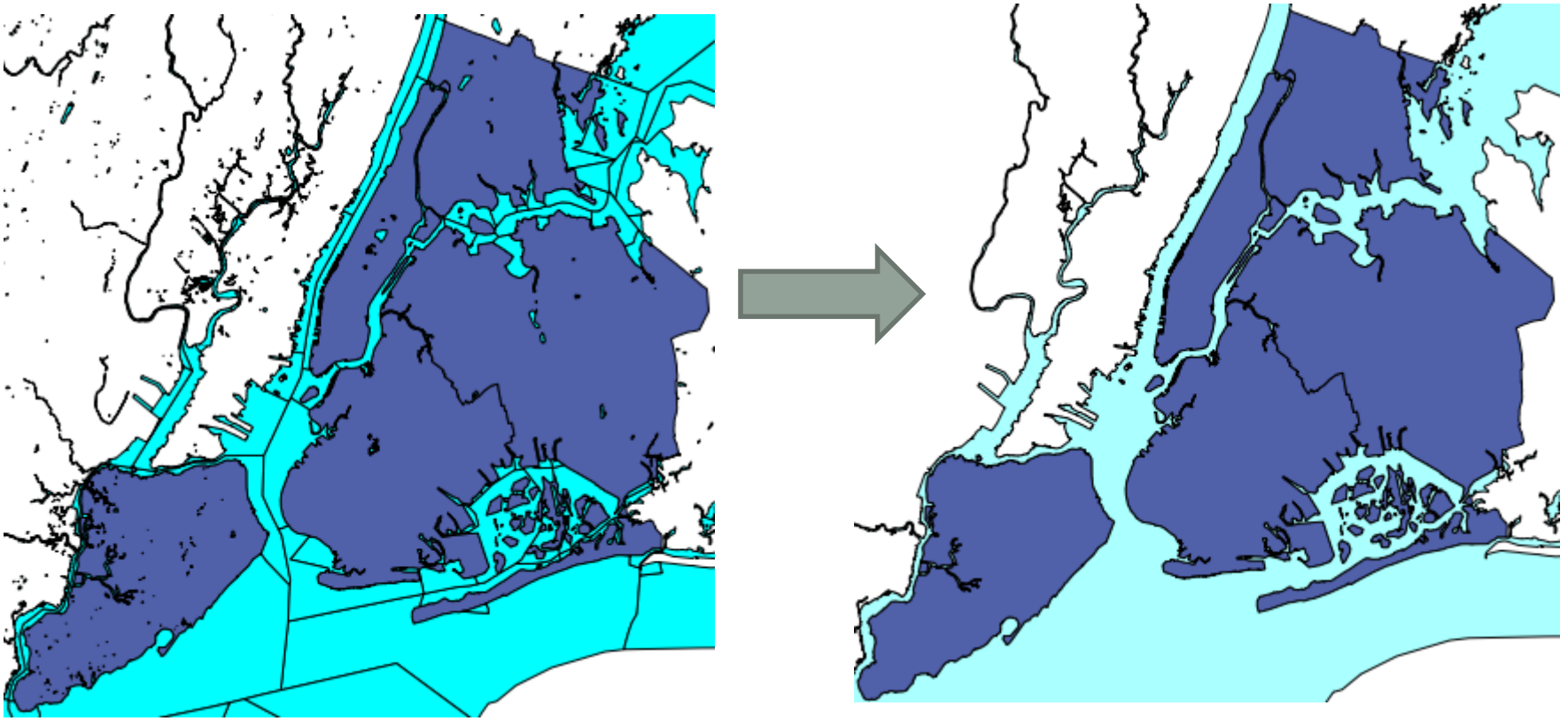


NAD83 **NY State Plane Long Island (ft)**

Geoprocessing



Generalization



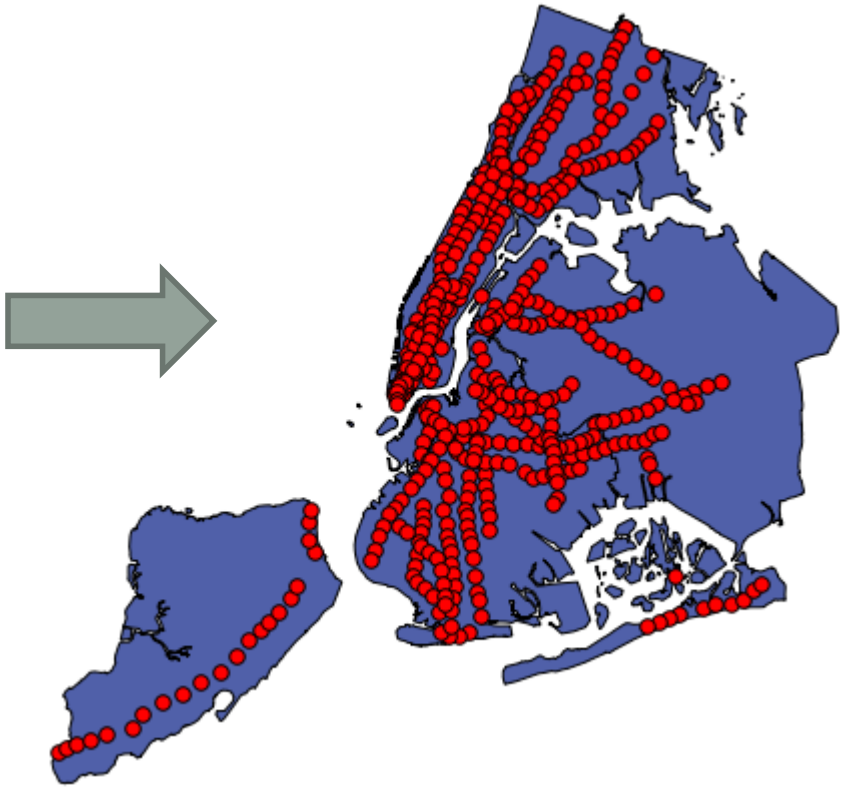
Generalized water (coastal and lakes) & landmarks (green space and facilities)

Transformation

stops_2ndcut.txt - Notepad

File Edit Format View Help

```
"stop_id","stop_name","stop_lat","stop_lon","train",
"L06","1 Av",40.730953,-73.981628,"L",
"119","103 St",40.799446,-73.968379,"1",
"624","103 St",40.7906,-73.947478,"6",
"A18","103 St",40.796092,-73.961454,"B C",
"706","103 St - Corona Plaza",40.749865,-73.8627,"7",
"A63","104 St",40.681711,-73.837683,"A",
"J14","104 St",40.695178,-73.84433,"J Z",
"623","110 St",40.79502,-73.94425,"6",
"705","111 St",40.75173,-73.855334,"7",
"A64","111 St",40.684331,-73.832163,"A",
"J13","111 St",40.697418,-73.836345,"J",
"226","116 St",40.802098,-73.949625,"2 3",
"622","116 St",40.798629,-73.941617,"6",
"A16","116 St",40.805085,-73.954882,"B C",
"117","116 St - Columbia University",40.807722,-
73.96411,"1",
"J12","121 St",40.700492,-73.828294,"J Z",
"116","125 St",40.815581,-73.958372,"1",
"225","125 St",40.807754,-73.945495,"2 3",
"621","125 St",40.804138,-73.937594,"4 5 6",
"A15","125 St",40.811109,-73.952343,"A B C D",
"224","135 St",40.814229,-73.94077,"2 3",
"A14","135 St",40.817894,-73.947649,"B C",
"115","137 St - City College",40.822008,-
```



NYC Facilities (schools, libraries, hospitals) & MTA (subway and train stations)

Database Features and Tables

- “A” Objects: geographic features used for mapping data or representing landmarks (2010 Census TIGER files, NYC Facilities database, MTA transit features)
- “B” Objects: data tables that can be joined to A features for mapping data (2010 Census, 5-year American Community Survey, ZIP Code Business Patterns)
- “C” Objects: represent the actual legal or statistical boundaries of some A objects, for reference only
- “X” Objects: represent the full, original features that A Objects were created from (for users who wish to add additional detail)

NYC Geodatabase (nyc_gdb)

P The NYC Geodatabase (nyc_gdb) is a resource designed 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.

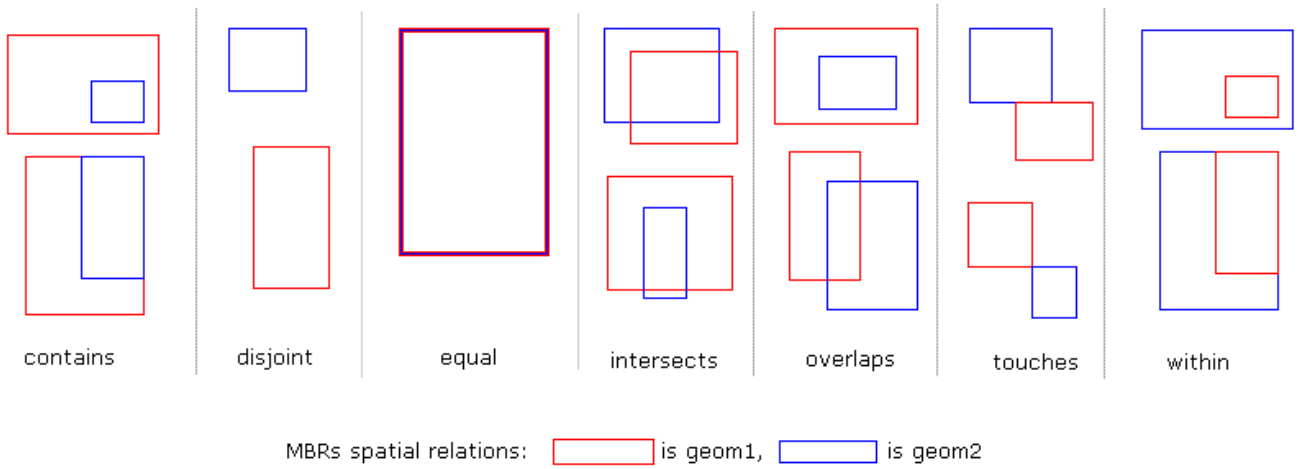
The contents of the databases are identical and include geographic features and statistical areas from the US Census Bureau, transit and public facility point features from the City, and census data at the PUMA, ZIP / ZCTA, and census tract levels. All features share a common coordinate system: NAD 83 NY State Plan Long Island (feet).

The databases will be updated bi-annually each winter and summer; the current version is **nyc_gdb_july2013**. *NOTE: when you download a database the documentation files are included in the ZIP file*

Name	File Type	Size	Download
nyc_gdb for QGIS (Spatialite version)	sqlite	29 MB	nyc_gdb_july2013a.zip
Introduction and Tutorial (Spatialite version)	pdf	831 KB	intro_nycgdb_foss.pdf
nyc_gdb for ArcGIS (MS Access version)	mdb	29 MB	nyc_gdb_july2013b.zip
Introduction and Tutorial (MS Access version)	pdf	378 KB	intro_nycgdb_arc.pdf
Data Dictionary	pdf	171 KB	datadiction_july2013.pdf
Metadata	xml	4 KB	nyc_gdb_july2013.xml
Archive of previous versions			

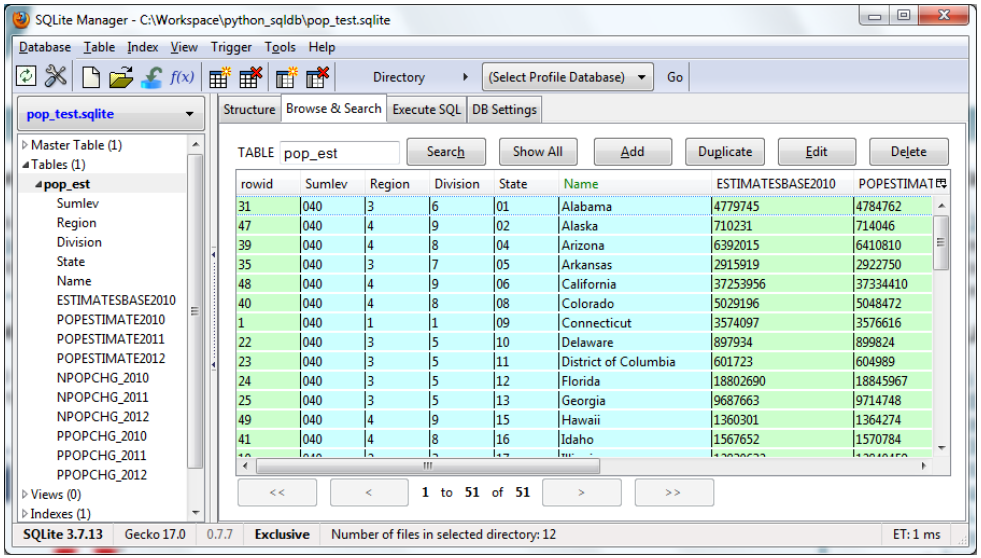
Why Spatial Databases?

- Organization and storage: keep related geographic features and data tables together in one place
- Harness benefits of relational databases
- Speed processing and analysis by connecting programming languages to DB
- In the case of the open source DBs, extend the power of open source GIS

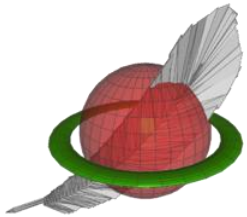


SQLite

- Embedded Database: lightweight file-based db designed to be seamlessly embedded in applications
- Created by Dr. Richard Hipp in 2000
- Public domain and widely adopted - http://sqlite.org/
- Like many DBs the interface is command line, but several front ends have been developed – SQLite Manager



Spatialite



- An extension of SQLite that can hold geographic objects and perform Spatial SQL functions, compliant with open standards (OGC - similar to PostGIS)
- Developed by Alessandro Furieri in 2008
- Open source: http://www.gaia-gis.it/gaia-sins/
- Command-line interface and GUI
- Also supported by QGIS via QGIS Browser and DB Manager
- Spatialite Cookbook: http://www.gaia-gis.it/gaia-sins/spatialite-cookbook/

Spatial Databases and Geometry

- DBs that have been optimized to store spatial data and perform geographic functions
- Geographic vector features stored as series of coordinates tied to a specific CRS

Geometry Type	WKT example
POINT	**POINT(123.45 543.21)**
LINESTRING	**LINESTRING(100.0 200.0, 201.5 102.5, 1234.56 123.89)** *three vertices*
POLYGON	**POLYGON((101.23 171.82, 201.32 101.5, 215.7 201.953, 101.23 171.82))** *exterior ring, no interior rings* **POLYGON((10 10, 20 10, 20 20, 10 20, 10 10), (13 13, 17 13, 17 17, 13 17, 13 13))** *exterior ring, one interior ring*
MULTIPOINT	**MULTIPOINT(1234.56 6543.21, 1 2, 3 4, 65.21 124.78)** *three points*
MULTILINESTRING	**MULTILINESTRING((1 2, 3 4), (5 6, 7 8, 9 10), (11 12, 13 14))** *first and last linestrings have 2 vertices each one; the second linestring has 3 vertices*
MULTIPOLYGON	**MULTIPOLYGON(((0 0,10 20,30 40,0 0),(1 1,2 2,3 3,1 1)),** **((100 100,110 110,120 120,100 100)))** *two polygons: the first one has an interior ring*

Structured Query Language (SQL)

- SQL is the language for creating and manipulating relational databases; originally based on relational algebra, it uses declarative commands in English
- The structure of the relational database and SQL were designed to be independent of any specific hardware or software (Codd 1970)

```
SELECT zcta, bcode, HD01_S001 AS pop2010
FROM a_zctas, b_zctas_2010census
WHERE zcta=geoid2
AND bcode="36061" AND pop2010 > 200
ORDER BY pop2010
```

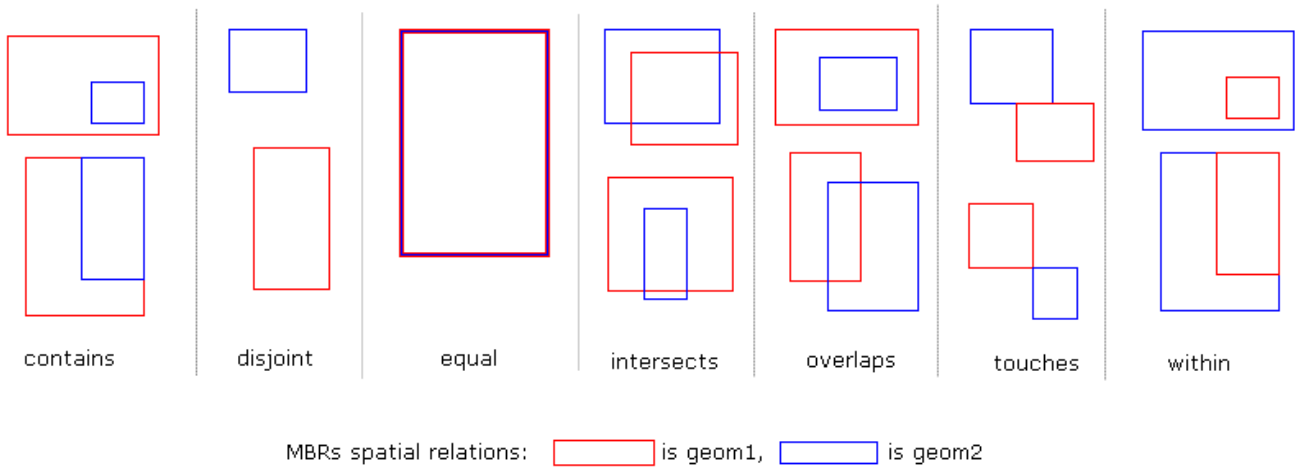
SQL and Spatial SQL (OGC)

```
SELECT stop_name, trains, ST_DISTANCE(a_zctas.geometry,
a_subway_stations.geometry) AS dist
FROM  a_zctas,  a_subway_stations
WHERE  zcta  = "10010" AND  dist <= 2640
ORDER  BY  dist
```

```
SELECT  stop_name,  trains
FROM  a_zctas,  a_subway_stations
WHERE  zcta  = "10010"
AND  ST_Within  (a_subway_stations.geometry,
a_zctas.geometry)
```

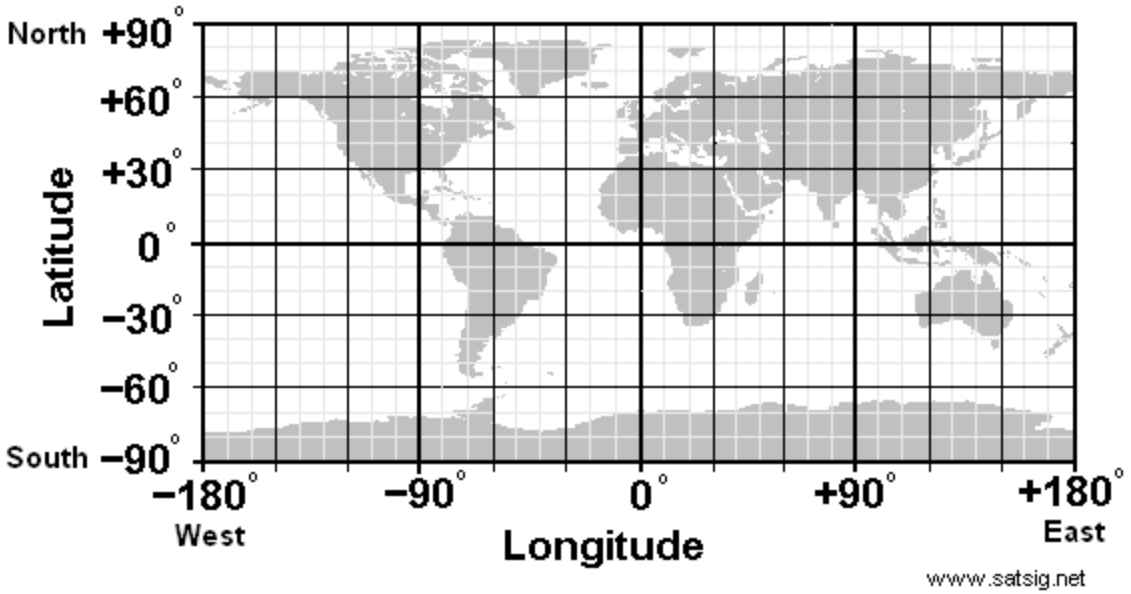
Examples

SpatiaLite 2.4.0 SQL functions reference list:
http://www.gaia-gis.it/spatialite-2.4.0/spatialite-sql-2.4.html



Also: Length, Perimeter, Area, Distance, Centroid, Envelope, Min X Y, Max X Y, Union, Buffer, Transform…

Long, Lat (X,Y)

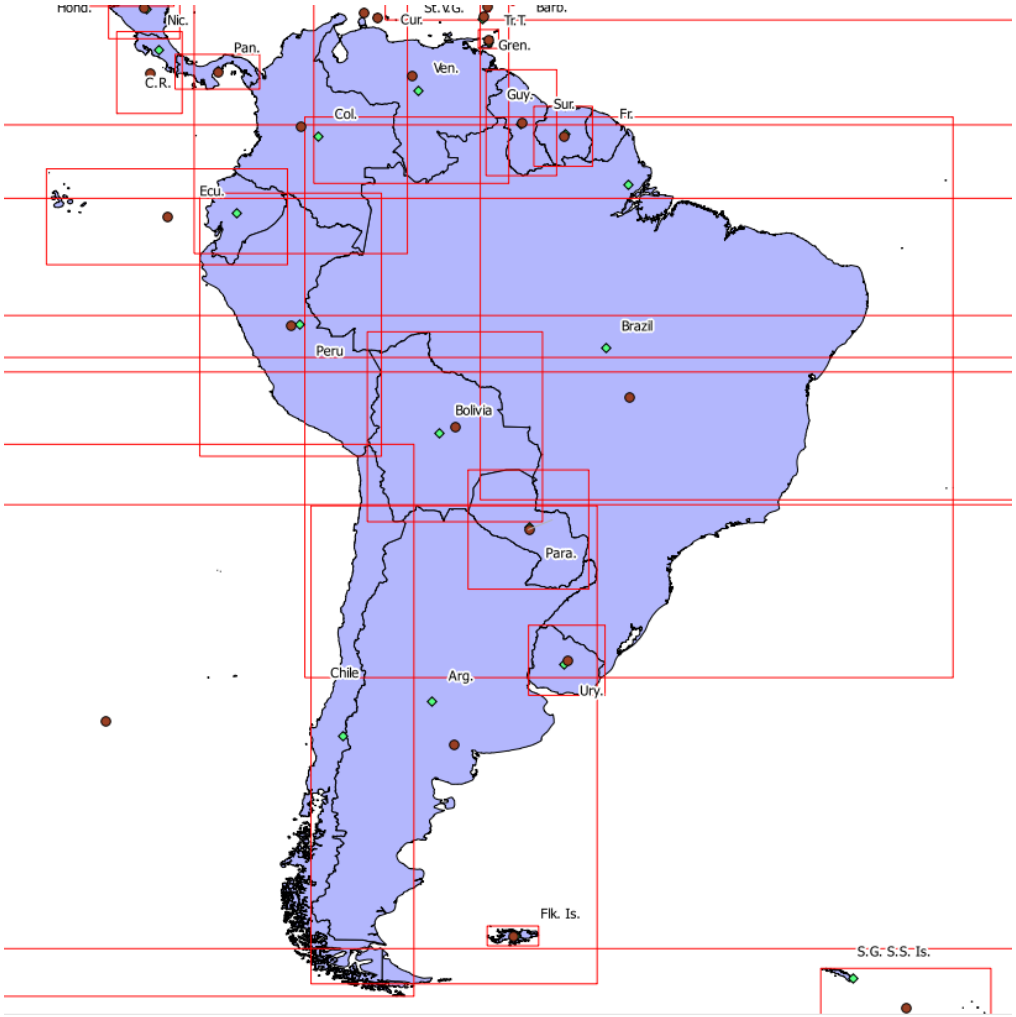


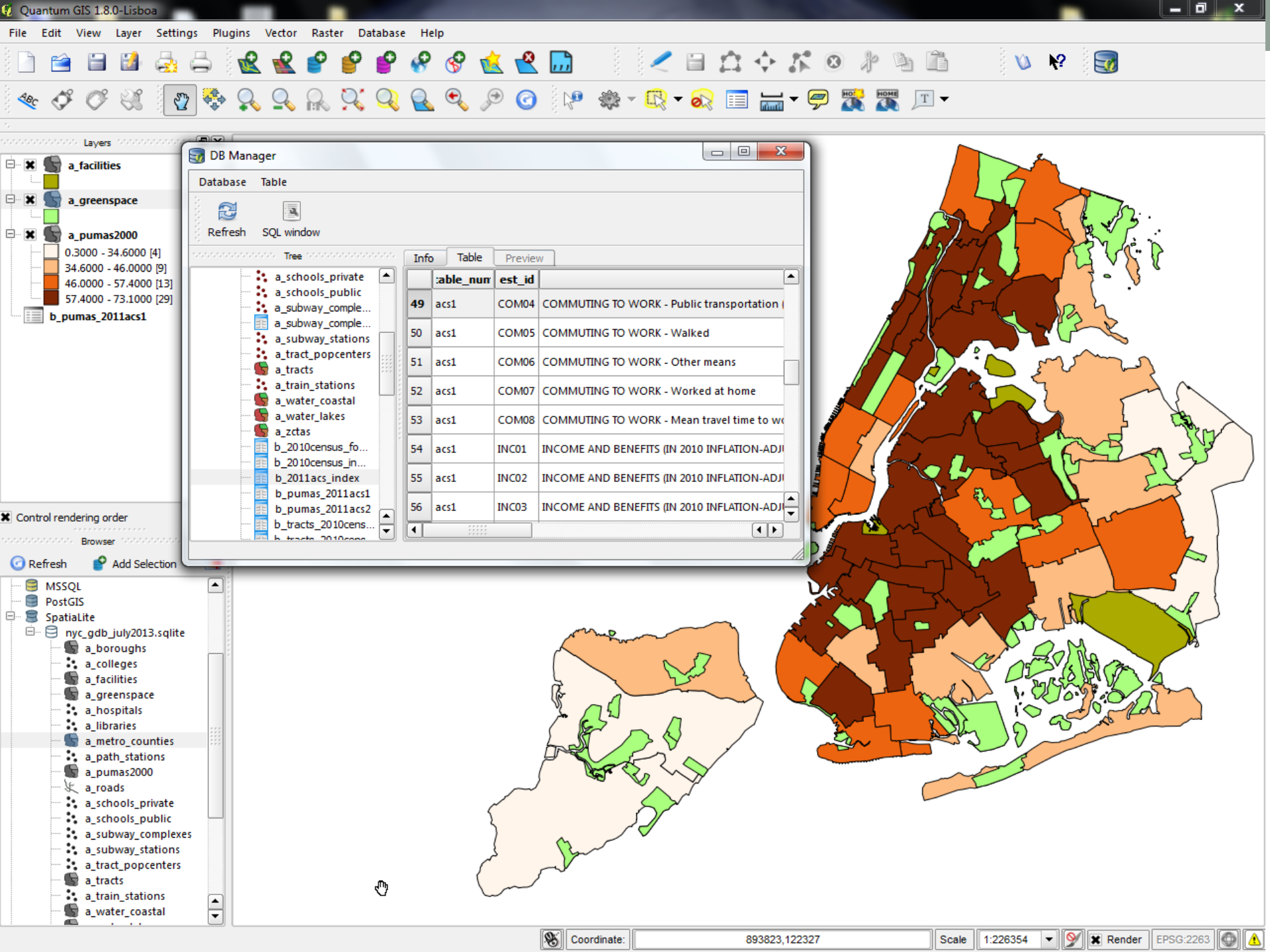
Y

X

Geometries and MBRs

- http://www.gaia-gis.it/gaia-sins/spatialite-cookbook/html/wkt-wkb.html




Quantum GIS 1.8.0-Lisboa
File Edit View Layer Settings Plugins Vector Raster Database Help
Layers
a_facilities
a_greenspace
a_pumas2000
0.3000 - 34.6000 [4]
34.6000 - 46.0000 [9]
46.0000 - 57.4000 [13]
57.4000 - 73.1000 [29]
b_pumas_2011acs1
Control rendering order
Browser
Refresh
Add Selection
MSSQL
PostGIS
SpatiaLite
nyc_gdb_july2013.sqlite
a_boroughs
a_colleges
a_facilities
a_greenspace
a_hospitals
a_libraries
a_metro_counties
a_path_stations
a_pumas2000
a_roads
a_schools_private
a_schools_public
a_subway_complexes
a_subway_stations
a_tract_popcenters
a_tracts
a_train_stations
a_water_coastal
DB Manager
Database Table
Refresh
SQL window
Tree
a_schools_private
a_schools_public
a_subway_comple...
a_subway_comple...
a_subway_stations
a_tract_popcenters
a_tracts
a_train_stations
a_water_coastal
a_water_lakes
a_zctas
b_2010census_fo...
b_2010census_in...
b_2011acs_index
b_pumas_2011acs1
b_pumas_2011acs2
b_tracts_2010cens...
Info
Table
Preview
:able_num
est_id
49 acs1 COM04 COMMUTING TO WORK - Public transportation
50 acs1 COM05 COMMUTING TO WORK - Walked
51 acs1 COM06 COMMUTING TO WORK - Other means
52 acs1 COM07 COMMUTING TO WORK - Worked at home
53 acs1 COM08 COMMUTING TO WORK - Mean travel time to w
54 acs1 INC01 INCOME AND BENEFITS (IN 2010 INFLATION-ADJ
55 acs1 INC02 INCOME AND BENEFITS (IN 2010 INFLATION-ADJ
56 acs1 INC03 INCOME AND BENEFITS (IN 2010 INFLATION-ADJ
Coordinate: 893823,122327
Scale 1:226354
Render
EPSG:2263

Limitations

- When creating a new table from a query (`CREATE TABLE AS SELECT…`), you cannot specify a primary key or data types for calculated fields (limitation of SQLite)
- When importing tables or shapefiles, you cannot specify keys, constraints, or data types (limitation of Spatialite GUI)
- SQLite's ALTER statement is very limited; you cannot delete columns from tables or change data types

- Coping method: Create a new, blank table that has the structure you want, copy data from the existing table into the blank table

Typical Procedures for Creating Well-formed data tables (post-import)

```
CREATE  TABLE  newtable (
newid TEXT  NOT  NULL  PRIMARY  KEY,
otherid TEXT,
value1  INTEGER,
value2  REAL)

SELECT AddGeometryColumn ( "newtable", "geometry", 2263,
"TYPE OF GEOMETRY", "XY")

INSERT INTO newtable (newid, otherid, value1, value2,
geometry)
SELECT  shapeid, label, popvar, housevar, geometry
FROM  shapefile

DROP  shapefile
```

When do you use Sqlite / Spatialite?

- Small to medium size projects
- Couple of users
- Portability is important
- Open source environment

- Not for large projects or datasets
- Not for networked environments where multiple writers and readers required
- Does not interact with proprietary GIS

THE NYC GEODATABASE

An Open Source Spatialite Database

http://www.baruch.cuny.edu/geoportal/nyc_gdb/

Frank Donnelly, Geospatial Data Librarian

Baruch College CUNY

francis.donnelly@baruch.cuny.edu