

# Ocean State Spatial Database Documentation

Frank Donnelly, Head of GIS & Data Services, Brown University Library

May 1, 2024

[https://github.com/Brown-University-Library/geodata\\_ossdb](https://github.com/Brown-University-Library/geodata_ossdb)

---

## Introduction

The Ocean State Spatial Database (OSSDB) is a geodatabase for conducting basic geographic analysis and thematic mapping within the State of Rhode Island. Created by the Brown University Library's GeoData@SciLi team, it is intended to serve as a basic foundation for contemporary mapping projects, and as an educational tool for supporting GIS coursework and introducing spatial databases and SQL. It includes geographic features and statistical areas from the US Census Bureau, census data tables, and public facility features generated from various state agencies. All features share a common coordinate system: NAD 83 RI State Plane (US-feet), EPSG 3438. Each table has a prefix that indicates what type of information it contains.


This document provides an overview of the database content and structure, with illustrations of how to use the database with different software packages. The primary database is in the open source SQLite / Spatialite format, which can be used with many desktop GIS and relational database software packages, including QGIS and ArcGIS Pro. A copy of the database is also available in the ESRI file geodatabase format, as an alternative for ArcGIS users.

## Contents

<b>1</b>	<b>Rights and Use</b>	<b>2</b>
<b>2</b>	<b>Database Features and Updates</b>	<b>2</b>
<b>3</b>	<b>Database</b>	<b>4</b>
	Format . . . . .	4
	Contents . . . . .	4
<b>4</b>	<b>Adding Layers to ArcGIS Pro</b>	<b>7</b>
<b>5</b>	<b>Adding Layers to QGIS</b>	<b>9</b>
<b>6</b>	<b>Using the QGIS DB Manager</b>	<b>10</b>
<b>7</b>	<b>Database Software</b>	<b>11</b>
	Spatialite GUI . . . . .	12
	DB Browser for SQLite . . . . .	12

# 1 Rights and Use

Disclaimer: Every effort was made to ensure that the data, which was compiled from public sources, was processed and presented accurately. The creators and Brown University 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 and fitness of the data for their purposes.

The database and associated documentation are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International 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.

The database is appropriate for thematic and reference mapping at a state and county level, and for thematic mapping at a municipal level. While it can be used for creating reference maps below the county 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 and the number of water features selected for inclusion). Users will have to judge for themselves based on their intended purpose.

## 2 Database Features and Updates

The 2nd version of the database was released in May 2024, and replaces the 1st version released in July 2023. Changes between this version and the last:

1. All of the c census ACS tables for census tracts, county subdivisions, and ZCTAs were updated, replacing data from 2017-2021 with 2018-2022.
2. New c tables for the 2020 Census Demographic Profile were added for census tracts, county subdivisions, and ZCTAs, along with a lookup table to associate variable IDs with names. There are two tables for each geography, one with counts and another with percent totals.
3. Old c tables from the 2020 Census Public Redistricting File were dropped, as the new profile tables are more comprehensive.
4. New c tables from the 2021 Census ZIP Code Business Patterns were added for ZCTAs. There are two tables; an employment table with a count of business establishments, employment, and wages, and an industry table with counts of establishments by NAICS 2-digit sector codes. Also included is a lookup table for NAICS codes and a cross walk table that was used to aggregate ZIP Code data to the ZCTA level.
5. d\_public\_libraries was updated, replacing data from 2021 with data from 2022.
6. d\_colleges and d\_schools\_pk12 were updated, replacing data from 2023 with data from 2024.
7. The new d\_colleges feature set only includes colleges and universities, and no longer includes post-secondary institutions such as technical institutes, beauty academies, and truck driving schools. The data source for these features was not reliable, so we dropped them.

**Version: ossdb\_2024\_05**

Updates for this version are highlighted

Table Name	Description	Date
a_census_tract_popcenters	Population centroids for census tracts	2020
a_census_tracts	Census tracts	2021
a_counties	Counties	2021
a_county_subdivs	County subdivisions (municipalities)	2021
a_state	State	2021
a_zctas	ZIP code tabulation areas	2021
b_census_tract_bndy	Census tract boundaries	2021
b_county_bndy	County boundaries	2021
b_county_subdiv_bndy	County subdivision boundaries	2021
b_state_bndy	State boundary	2021
b_zcta_bndy	ZIP code tabulation area boundaries	2021
c_acs2022_lookup	Variable lookup table for American Community Survey	2018-2022
c_census_2020_lookup	Variable lookup table for 2020 Census data	2020
c_subdivs_acs2022_pophous	ACS population and housing data for county subdivisions	2018-2022
c_subdivs_acs2022_socoecon	ACS social and economic data for county subdivisions	2018-2022
c_subdivs_census2020_count	2020 Census counts for county subdivisions	2020
c_subdivs_census2020_pct	2020 Census percent totals for county subdivisions	2020
c_tracts_acs2022_pophous	ACS population and housing data for census tracts	2018-2022
c_tracts_acs2022_socoecon	ACS social and economic data for census tracts	2018-2022
c_tracts_census2020_count	2020 Census counts for census tracts	2020
c_tracts_census2020_pct	2020 Census percent totals for census tracts	2020
c_zbp2021_indcodes	ZBP NAICS 2-digit sector codes	2021
c_zctas_acs2022_pophous	ACS population and housing data for ZCTAs	2018-2022
c_zctas_acs2022_socoecon	ACS social and economic data for ZCTAs	2018-2022
c_zctas_census2020_count	2020 Census counts for ZCTAs	2020
c_zctas_census2020_pct	2020 Census percent totals for ZCTAs	2020
c_zctas_zbp2021_emp	Business Patterns establishments and employment for ZCTAs	2021
c_zctas_census2020_ind	Business Patterns establishments by industry for ZCTAs	2021
c_zip2zcta_2022	ZIP Code to ZCTA cross walk	2022
d_airports	Major airports	2023
d_coastal_water	Coastal water	2021
d_colleges	Colleges and universities	2024
d_hospitals	Hospitals	2023
d_lakes_rivers	Major lakes and rivers	2021
d_narragansett_resv	Narragansett Indian Reservation	2021
d_public_libraries	Public libraries	2022
d_rail_stations	Passenger railroad stations	2023
d_rails	Passenger railroad tracks	2021
d_roads	Roads and highways	2021
d_schools_pk12	Pre-K to 12th grade schools	2024
x_all_water	Complete water layer from which other water layers were created	2021
z_metadata	Metadata table for Ocean State Spatial Database	2024

Consult the external metadata files included in the metadata folder accompanying each download for details about the features and tables. Records are expressed in the OpenGeoMetadata Aardvark schema and are stored in plain-text yaml and markdown formats.

## 3 Database

### Format

A spatial database is a relational database that has been enhanced to hold spatial objects or geographic features. The OSSDB is in an open source format called Spatialite. Built on the popular SQLite database, it consists of a single .sqlite file that can be used in both QGIS and ArcGIS Pro, as well as database applications like the QGIS DB Manager, Spatialite GUI, and DB Browser for SQLite. A copy of the database is also available in the ESRI file geodatabase format, as an alternative for ArcGIS users.

Spatial databases are better for organizing and structuring data and projects relative to shapefiles and individual data tables, as multiple features and attributes can be stored in a single database. Spatialite has the added advantage of allowing users to write SQL and spatial SQL queries against features and tables, in addition to using the stand-alone tools that desktop GIS software provides. As a file-based database, Spatialite dbs can easily be copied and distributed, but are not designed for multiple, simultaneous users.

### Contents

Objects in the database are categorized and named with a prefix to differentiate different types of features. A brief list of the database objects is provided below; for full details consult the metadata records for individual objects. All geospatial features in the database are projected in the NAD 83 RI State Plane (US-feet) coordinate reference system, ID EPSG 3438.

#### A Tables

Objects that begin with the prefix "a" are geographic features that represent land areas for thematic mapping, or points for distance measurement. They include legal (state, county, county-subdivision) and statistical (census tracts, ZIP Code Tabulation Areas) areas delineated by the Census Bureau based on the 2020 census and drawn from the 2021 Census TIGER Line Files. Area features have been clipped using the d\_coastal\_water layer to remove large areas of water. Each "a" layer contains both a long and short version of the Census Bureau's ANSI / FIPS code, which uniquely identifies each feature and can be used for joining tabular census data to the features.

**a\_census\_tract\_popcenters** : Population centroids for census tracts based on the 2020 Census. Represents the center of the population's distribution within each tract.

**a\_census\_tracts** : 2020 census tracts; census statistical areas designed to have an ideal size of 4,000 residents, with a range of 1,200 to 8,000.

**a\_counties** : 2020 counties. Counties in Rhode Island have no legal or governmental authority, but are used for statistical purposes.

**a\_county\_subdivs** : 2020 county subdivisions, which in Rhode Island consist of municipalities (cities and towns).

**a\_state** : 2020 state land area.

**a\_zctas** : 2020 ZIP Code Tabulation Areas; census statistical areas created by aggregating census blocks based on postal addresses, to create geographic approximations of USPS ZIP Codes.

## **B Tables**

Objects that begin with the prefix "b" are geographic features that represent boundaries for reference mapping. They are the original, unclipped versions of the "a" polygon features, with areas that incorporate both land and water. Each layer contains a short version of the Census Bureau's ANSI / FIPS code, which uniquely identifies each feature.

**b\_census\_tract\_bndy** : 2020 census tract boundaries.

**b\_county\_bndy** : 2020 county boundaries.

**b\_county\_subdiv\_bndy** : 2020 county subdivision boundaries.

**b\_state** : 2020 state boundary.

**b\_zctas** : 2020 ZIP Code Tabulation Area boundaries.

## **C Tables**

Objects that begin with the prefix "c" are non-spatial data tables from the US Census, with data reported as values and percentages. These tables can be joined to geographic "a" features so that quantities can be mapped and evaluated spatially. Most "c" tables contains both a long and short version of the Census Bureau's ANSI / FIPS code, which uniquely identifies each record and can be used for joining tabular census data to the "a" features.

Tables are named based on their geography, dataset, and year. Column names are codes that uniquely identify each variable. For each dataset there is an index table named with the suffix "lookup", that relates column codes to variable names.

### American Community Survey

The American Community Survey (ACS) is an ongoing sample survey of the population. The series in the database are from the 5-year period estimates; the year indicates the year of release and final year of the estimate range (i.e. 2022 represents 2018-2022 5-year data). The values are published as estimates with a 90% confidence interval and margins of error (+/-). There are two data tables for each geography that represent a subset of the four demographic profiles (tables DP02 through DP05). Each individual variable is named based on its subject and consists of four adjacent

columns in this order: the estimate itself (identified by the suffix "e"), a margin of error for the estimate (suffix "m"), a percent total (suffix "pe"), and a margin of error for the percent total (suffix "pm").

- c\_YEARacs\_lookup
- c\_csubdivs\_YEAR\_pophous
- c\_tracts\_YEAR\_pophous
- c\_zctas\_YEAR\_pophous
- c\_csubdivs\_YEAR\_socecon
- c\_tracts\_YEAR\_socecon
- c\_zctas\_YEAR\_socecon

### Decennial Census

The decennial census is a 100% count of the population taken on April 1st. The decennial data in the database comes from the 2020 Demographic Profile table DP1, which provides a range of fundamental counts: population by age, sex, race, and ethnicity, households, families, and group quarters, and housing units by vacancy and tenure. There are two tables for each geography: one with counts, and another with percent totals. The lookup table correlates the column headings created by the Census with the variable names.

- c\_census2020\_lookup
- c\_csubdivs\_census2020\_count
- c\_tracts\_census2020\_count
- c\_zctas\_census2020\_count
- c\_csubdivs\_census2020\_pct
- c\_tracts\_census2020\_pct
- c\_zctas\_census2020\_pct

### Business Patterns

The Census Business Patterns is an annually produced administrative dataset that counts business establishments by geography. Data from the ZIP Code Business Patterns (ZBP) was used to generate data for ZCTAs; some ZIP Codes are point-based and do not have a corresponding areal ZCTA, so we aggregated data for these point-based ZIPs into the ZCTAs that contained them. There are two tables: one with a count of establishments, employment, and wages (1st quarter and annual, in millions of dollars), and another that counts establishments by industry, using the 2-digit sector codes from the North American Industrial Classification System (NAICS). The industry table includes counts and percent totals. There is a lookup table to relate NAICS codes with names, and the relational ZIP Code to ZCTA table that was used for aggregating the data.

- c\_zctas\_zbpYEAR\_emp
- c\_zctas\_zbpYEAR\_ind
- c\_zbpYEAR\_indcodes
- c\_zip2zcta\_year

### **D Tables**

Objects that begin with the prefix "d" are geographic point, line, and area features generated from many different sources, to be used for thematic or reference mapping.

**d\_airports** : polygon layer of major airports, created from the Census TIGER landmarks file and supplemented by the OpenStreetMap.

**d\_coastal\_water** : polygon layer of major features created from the Census TIGER water files for each county. Used as the clip layer for creating the "a" features.

**d\_colleges** : point layer of colleges and universities, created by geocoding the RI Department of Education's Directory against the RI Department of Transportation's E911 Site Locator.

**d\_hospitals** : point layer created by geocoding the RI Department of Health's License List against the RI Department of Transportation's E911 Site Locator.

**d\_lakes\_rivers** : polygon layer of major rivers and lakes extracted from the Census TIGER water files.

**d\_narragansett\_resv** : polygon layer of the Narragansett Indian Reservation, extracted from the Census TIGER AINANH files.

**d\_public\_libraries** : point layer created by geocoding the Institute of Museum and Library Service's Public Library Survey outlets against the RI Department of Transportation's E911 Site Locator.

**d\_rail\_stations** : point layer of all passenger railroad stations, created from the OpenStreetMap and supplemented with information from Amtrak and the MBTA.

**d\_rails** : line layer of passenger railroad tracts, extracted from the Census TIGER railroad files and improved with editing.

**d\_roads** : line layer of highways and roads, extracted from the Census TIGER road files.

**d\_schools\_pk12** : point layer of all preschools, kindergartens, elementary schools, and high schools, created by geocoding the RI Department of Education's Directory against the RI Department of Transportation's E911 Site Locator.

### **Other Tables**

**x\_water** : the original census TIGER water polygon layer that was used to create the coastal and lakes / rivers water layers. Includes all water features in Rhode Island and surrounding areas.

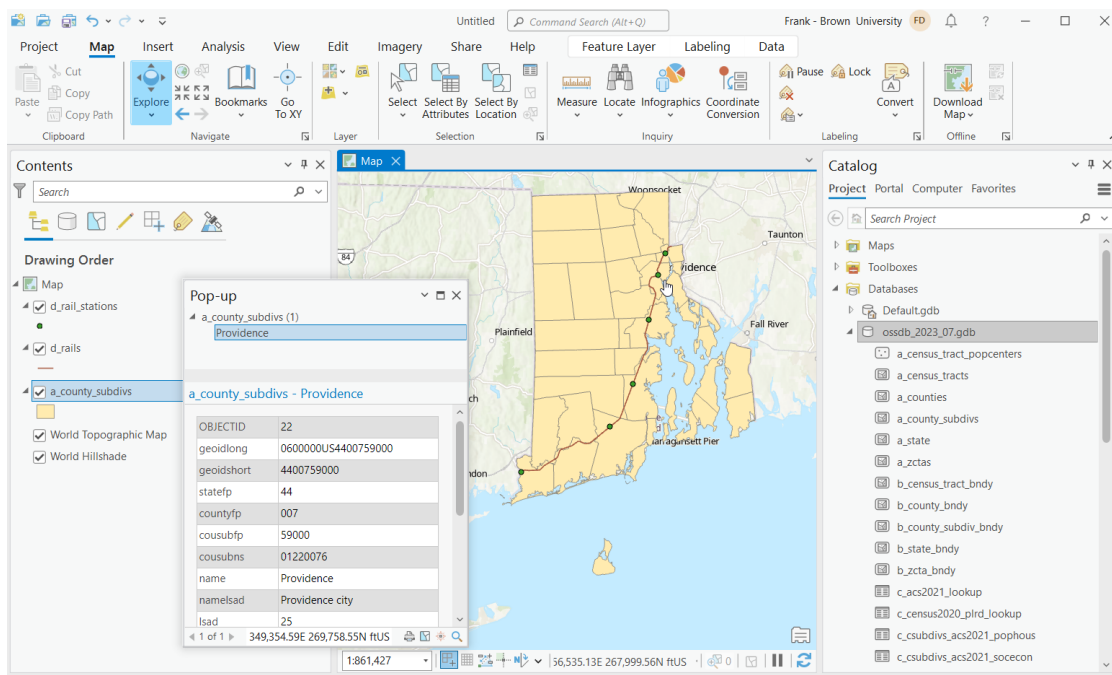
**z\_metadata** : a summary tables that lists all of the objects in the database, including information of sources and date last updated.

All other objects not enumerated here are core parts of the spatial database, designed to manage and support geographic features. These tables should not be removed or altered, otherwise the database could be rendered unusable.

## **4 Adding Layers to ArcGIS Pro**

ArcGIS Pro is proprietary desktop GIS software package published by ESRI that runs on MS Windows. ArcGIS users can use either the Spatialite format or the ESRI File Geodatabase format. In ArcGIS Pro 3.2, the Spatialite format is read-only, so you cannot edit or modify features or attributes in-place.

1. Download the OSSDB File Geodatabase, move it to a folder, and unzip the zip file. You may want to store the database in your user\_documents\_ArcGIS folder, as that's the default location for project files.
2. In ArcGIS Pro, use the *Catalog pane* (View - Catalog Pane), select the *Databases object*, right click and choose *Add Database*. Navigate to the folder where the database is stored, select it, and hit *OK* to establish a connection.
3. In the *Catalog pane*, hit the dropdown arrow beside the *Database object*, and then the dropdown arrow below the database to see the layers and tables.
4. Click on a layer or table to select it, hold down the left mouse button, and drag it into the *Map panel* to add it to your project.
5. You can overlay and symbolize the layers (right click on the layer in the *Contents pane* - *symbolology*), view attributes (right click on layer, choose *Attribute Table*) and use the analytical tools in the menus and toolbars just like you would any other vector layer.
6. To interpret the data in the "c" census data tables, add the lookup table to the project, and open it as an attribute table. This table correlates variable identifiers with variable names. To map the data, add the table and its matching geospatial "a" feature to a project. Select the "a" feature in the \*Contents pane\*, right click, and choose\* Joins and Relates - Joins\*. Join the "c" table to the "a" feature using the GEOIDLONG identifier that they share in common. Confirm the join, and then you can access the *Symbolology* of the "a" features to make a graduated area or symbol map from columns stored in the "c" table.



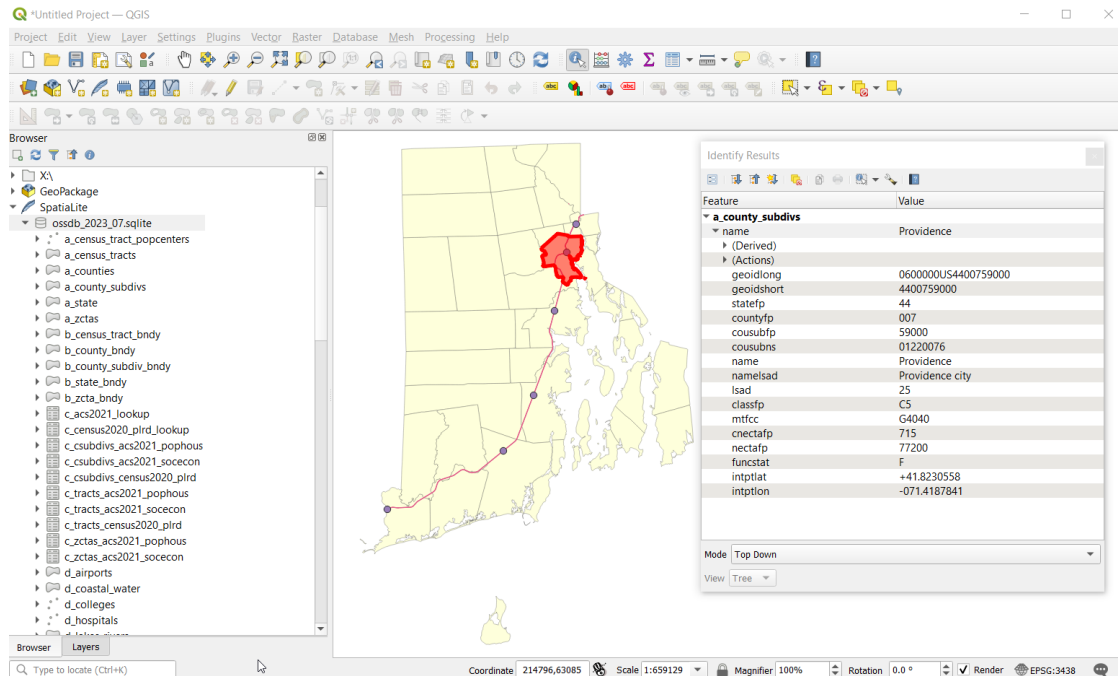


## 5 Adding Layers to QGIS

QGIS is a free and open source desktop GIS software package that runs on any operating system. New users should use the stable Long Term Release, not the Latest Release:

<https://www.qgis.org/>

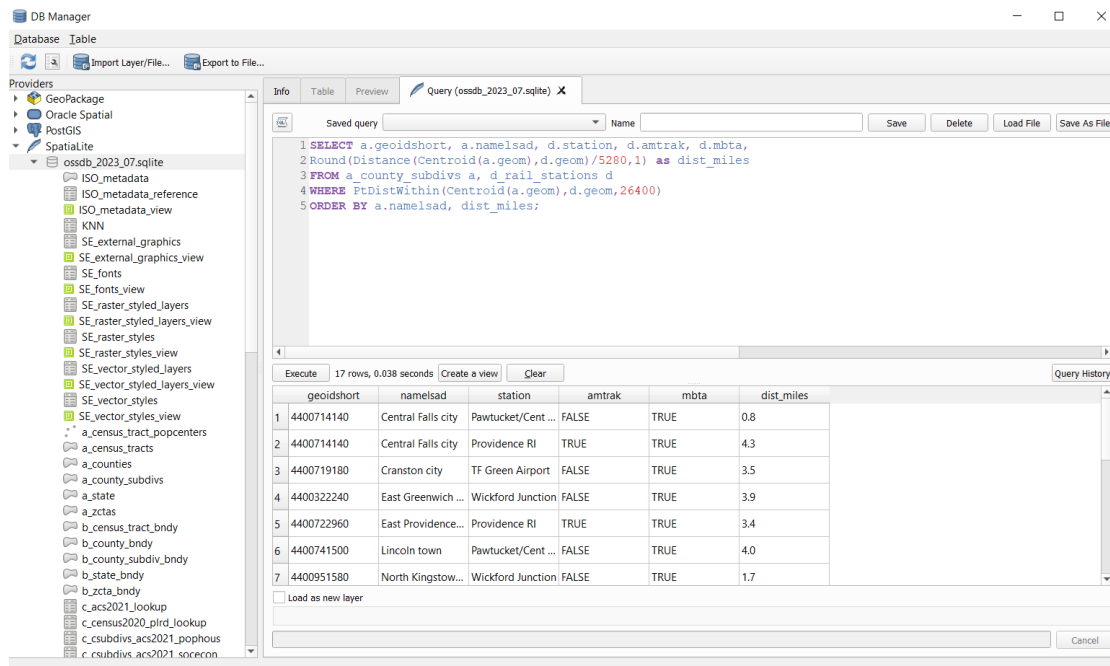
1. Download the OSSDB for SQLite file, move it to a folder, and unzip the zip file.
2. In the *QGIS Browser panel*, scroll down to the *Spatialite* option. Select it, right click, and choose *New Connection*. Browse to the folder where the *ossdb.sqlite* database is located, select it, and hit *Open* to establish a connection.
3. In the *Browser panel*, hit the dropdown arrow beside the *Spatialite* option, and then the dropdown arrow below the database to see the layers and tables.
4. Click on a layer or table to select it, hold down the left mouse button, and drag it into the map view to add it to your project.
5. You can overlay and symbolize the layers (right click on the layer in the *Layers panel - Properties - Symbolology*), view attributes (right click on layer, choose *Open Attribute Table*) and use the analytical tools in the menus and toolbars just like you would any other vector layer.
6. To interpret the data in the "c" census data tables, add the lookup table to the project, and open it as an attribute table. This table correlates variable identifiers with variable names. To map the data, add the table and its matching geospatial "a" feature to a project. Select the "a" feature in the *Layers panel*, right click, and choose *Properties*. On the *Joins tab*, hit the green plus symbol to add a new join. Join the "c" table to the "a" feature using the GEOIDLONG identifier that they share in common. Confirm the join, and then you can access the *Symbolology tab* of the "a" features to make a graduated area map from columns stored in the "c" table.



## 6 Using the QGIS DB Manager

The DB Manager in QGIS allows you to quickly preview features and tables without having to add them to a project. The SQL window allows you to execute SQL and spatial SQL queries, which provides more power and flexibility for processing and analyzing data.

1. Follow steps 1 and 2 in the previous section.
2. On the menu bar, choose *Database - DB Manager*.
3. In the *Providers menu* on the left, hit the dropdown arrow beside the *Spatialite* option, and the dropdown arrow beside the database to see the layers and tables. The objects that appear prior to the "a" features are internal system tables that you can ignore.
4. Click on a layer or table to select it, making it active. The window on the right will display metadata about the layer and how it was created (*Info tab*), the records in the attribute table (*Table tab*), and a preview of the feature's geometry (*Preview tab*).
5. You can add a layer or table to a project by selecting it in the *Providers menu*, right clicking, and choosing *Add to Canvas*.
6. The *SQL Window* button at the top will allow you to execute SQL and spatial SQL queries. Hit the button, and type your SQL query in the top window. Hit the *Execute* button to see the results in the bottom menu. You can load the result as a new, temporary layer to the map window to view it, and can save the result either as a view or a new table in the database.



For example, the following spatial SQL query calculates the distance (*Distance* function) from the geographic center (*Centroid* function) of every county subdivision (cities and towns) to railroad stations that are within (*PtDistWithin* function) 5 miles (expressed as 26,400 feet, as the units of the spatial reference system for the layers are in feet).

```
SELECT a.geoidshort, a.namesad, d.station, d.amtrak, d.mbt,
Round(Distance(Centroid(a.geom),d.geom)/5280,1) as dist_miles
FROM a_county_subdivs a, d_rail_stations d
WHERE PtDistWithin(Centroid(a.geom),d.geom,26400)
ORDER BY a.namesad, dist_miles;
```

For information about SQL and spatial SQL:

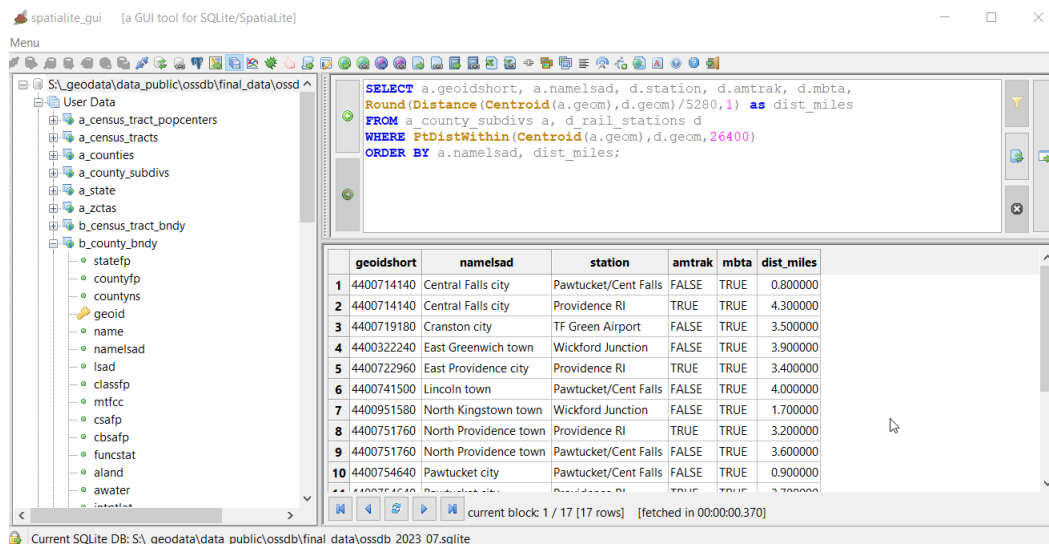
- SQL with SQLite: <https://www.sqlite.org/lang.html>
- Spatialite Functions:  
<https://www.gaia-gis.it/gaia-sins/spatialite-sql-5.0.1.html>

## 7 Database Software

There are a number of database viewers that can be used for writing SQL queries. These options are convenient for non-GIS users who prefer to work in a tabular environment, and for those who simply wish to export tables out of the database as files.

## Spatialite GUI

View tables and execute SQL and spatial SQL queries. The easiest way for Windows users to get started is to download the MS Windows binaries for the *previous stable version* from the bottom of the following page, and choose the 7zip file for the spatialite\_gui: <https://www.gaia-gis.it/gaia-sins/>. Mac users can use Homebrew [https://github.com/ryanpeek/spatiallite\\_demo](https://github.com/ryanpeek/spatiallite_demo), and Linux users can find the software in their package handler.



## DB Browser for SQLite

This app is easier to download and install than the Spatialite GUI, but is not customized for spatial data. It allows you to view tables and execute regular SQL queries, but spatial SQL is not supported (it is possible to add it as an extension with a bit of extra work). Available for all operating systems: <https://sqlitebrowser.org/>.