Query and Field Calculation

Lecture #12 | GEOG 510
GIS & Spatial Analysis in Public Health

Varun Goel

Outline

- Querying data
- Selecting features
- Attribute queries
- Spatial queries
- Calculate attributes
- Field statistics
- Field summary

Data Query

- Simplest form of spatial operation
- Can be *explicitly spatial* or operate only on *attributes* in a table
- Used to select and extract features from database that meet some set of conditions
- Can logically combine existing spatial features
- Supports higher level analytical operations, such as mathematical functions

Queries

- Simple and Compound
  - *Simple* queries use a single conditional statement
    - That is they specify one set of conditions a feature must meet to be selected
  - *Compound* queries use boolean logic to combine multiple statements

Queries

Simple selection:

records with Area > 20.0

ID	Area	Landuse	Municip
1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
5	22.1	Urban	City
6	30.2	Farm	Township
7	4.4	Urban	County

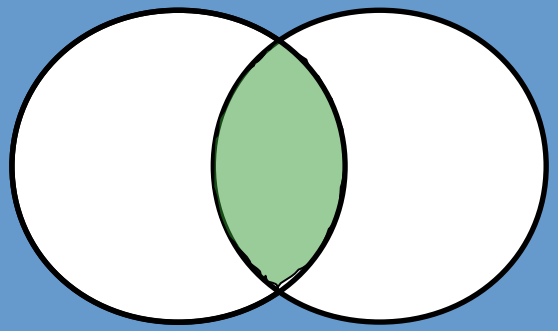
AND selection:

records with (Landuse = Urban) **and** (Municip = City)

ID	Area	Landuse	Municip
1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
5	22.1	Urban	City
6	30.2	Farm	Township
7	4.4	Urban	County

Boolean Logical Operations

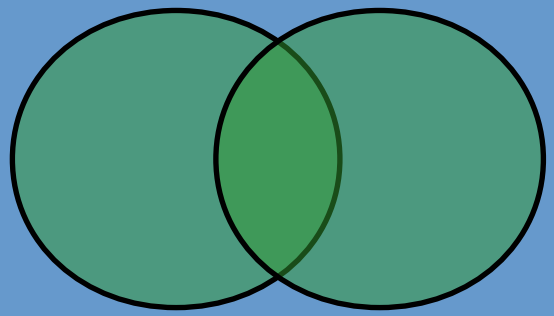
AND



Intersection

Identifies records that meet both of two conditions

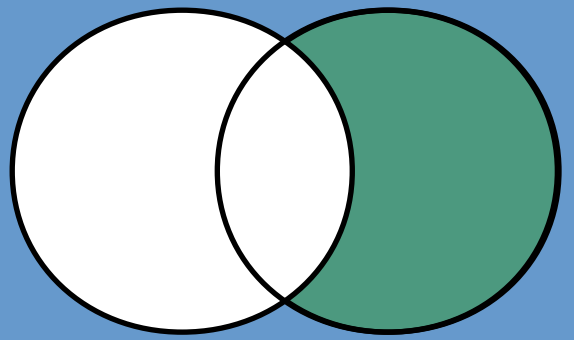
OR



Union

Identifies records that meet either of two conditions

NOT



Negation

Identifies records that do NOT meet one of the conditions

Boolean Logical Operations

Simple selection:

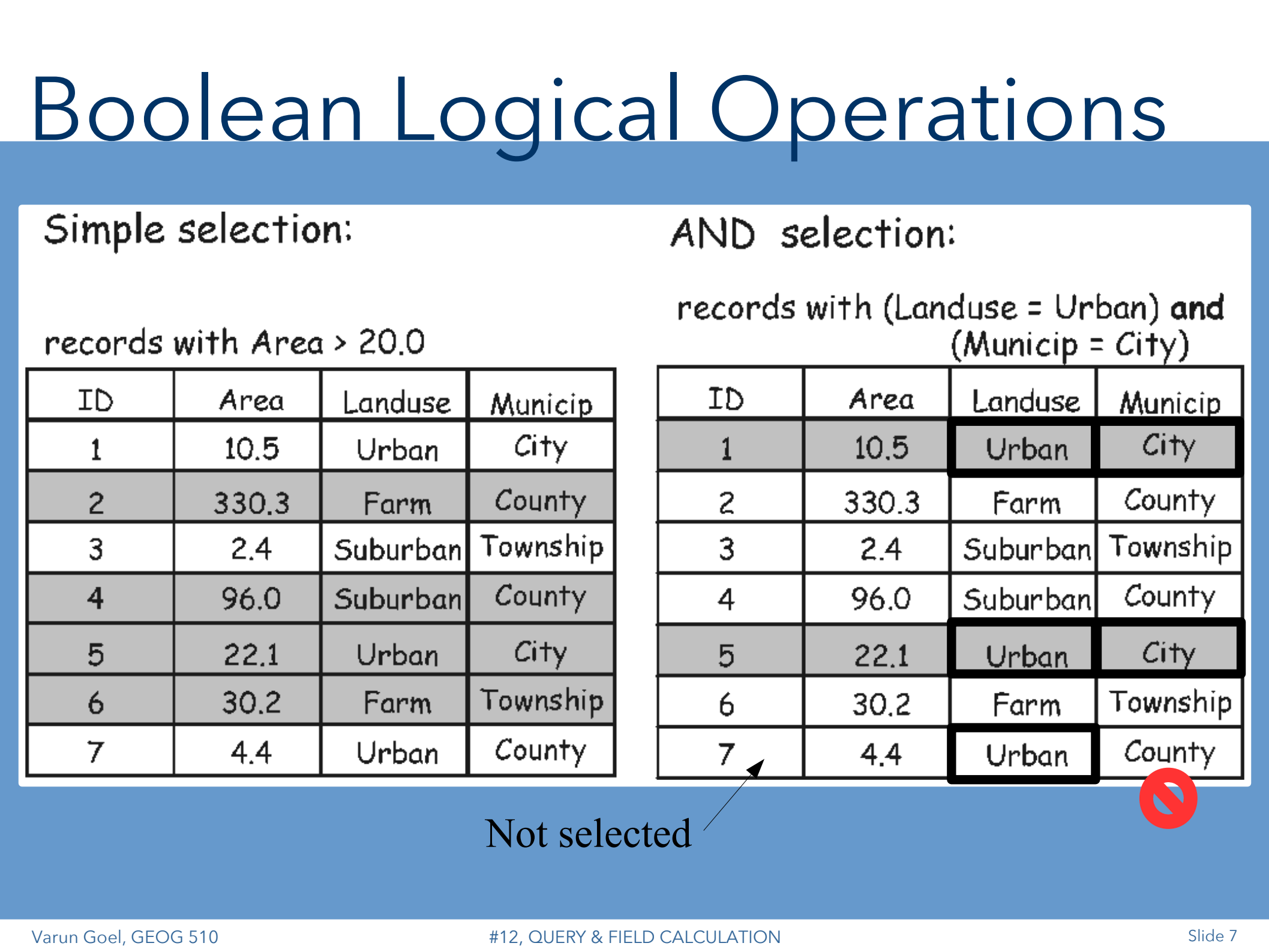
records with Area > 20.0

ID	Area	Landuse	Municip
1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
5	22.1	Urban	City
6	30.2	Farm	Township
7	4.4	Urban	County

AND selection:

records with (Landuse = Urban) **and** (Municip = City)

ID	Area	Landuse	Municip
1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
5	22.1	Urban	City
6	30.2	Farm	Township
7	4.4	Urban	County



Boolean Logical Operations

Simple selection:

records with Area > 20.0

ID	Area	Landuse	Municip
1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
5	22.1	Urban	City
6	30.2	Farm	Township
7	4.4	Urban	County

OR selection:

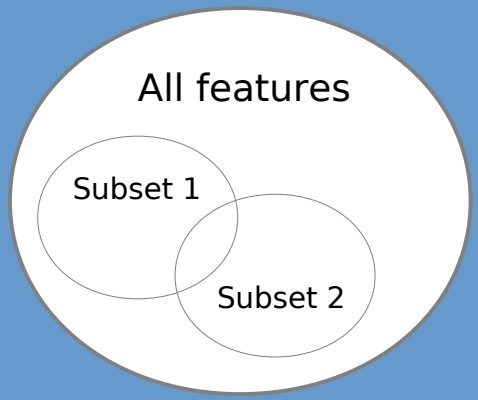
records with (Landuse = Urban) **or** (Municip = City)

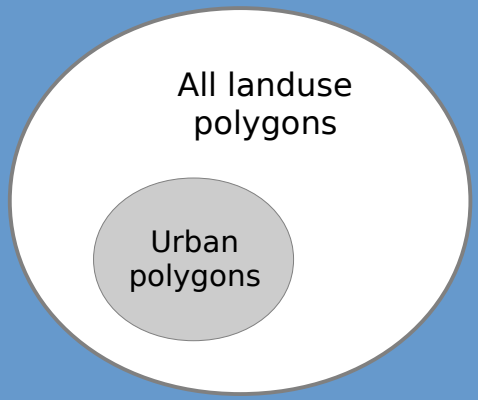
ID	Area	Landuse	Municip
* 1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
* 5	22.1	Urban	City
6	30.2	Farm	Township
* 7	4.4	Urban	County

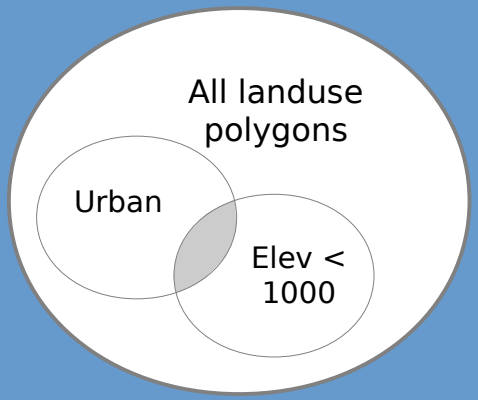
** All three selected*

Retrieving Data

- A set-based approach to managing features (or records) in database







- Select subset of records based on a logical expression
  - Can be simply equivalence expressions (`sel landuse = urban`)
  - Can be arithmetic expressions on quantities (`sel elevation < 1000`)
  - Can combine multiple conditions
    - (`sel landuse = urban AND elevation < 1000`)

Structured Query Language

- SQL
  - Provides a logical language for accessing data stored in a database
- Queries take the form:

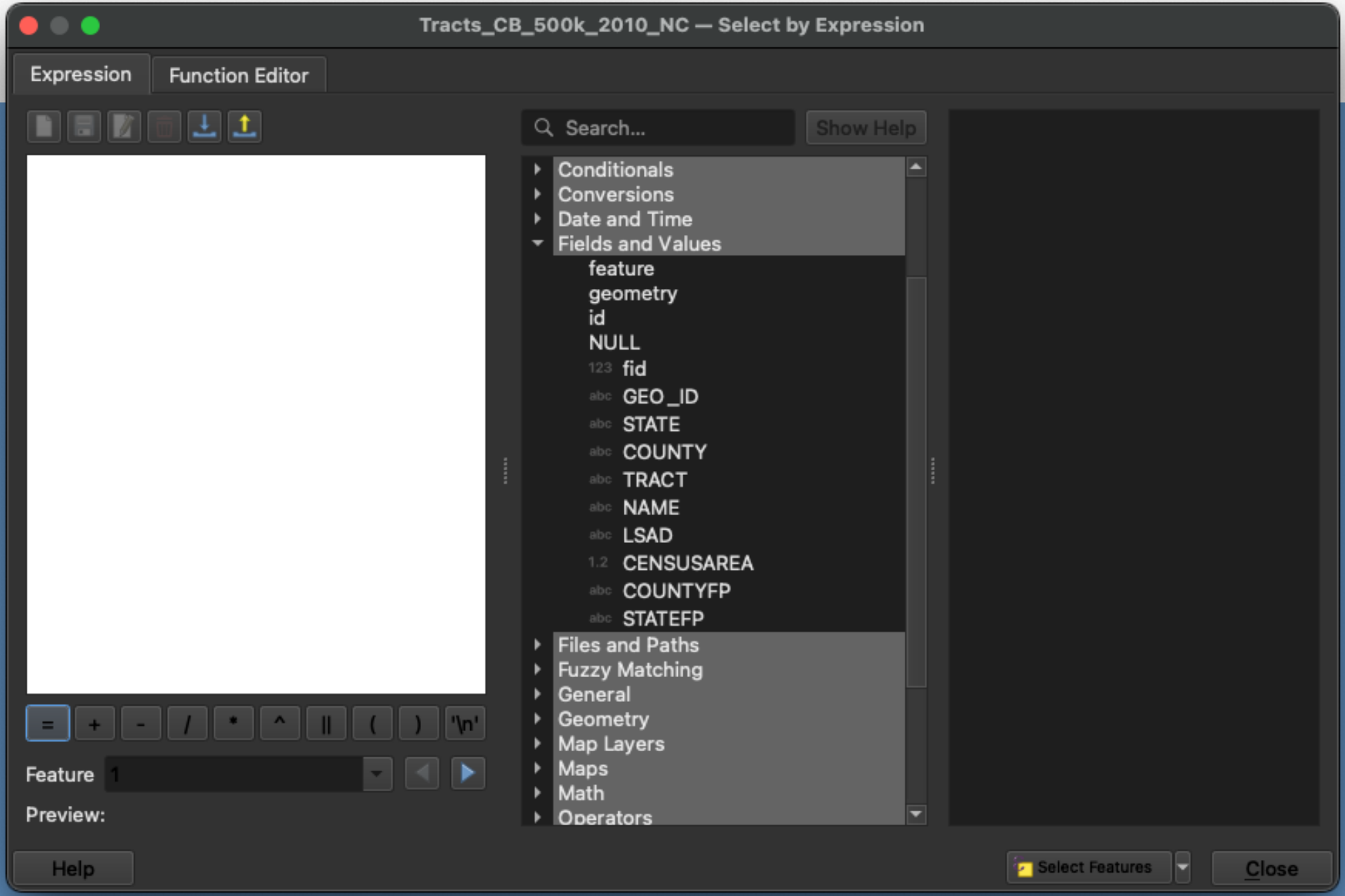
```
Select ____ from ____ where _____ AND/OR/NOT _____
```

- For example:

```
Select POLYGONS from STATES where AREA > 20.0
Select POLYGONS from COUNTIES where AREA > 5.0 AND POP > 2500
```

Selecting Features

- Select features by *attributes*
  - Graphically builds logical expressions to interact with attribute tables and relationship joins between them
  - Creates structured query languages (SQL) statements used with most database software

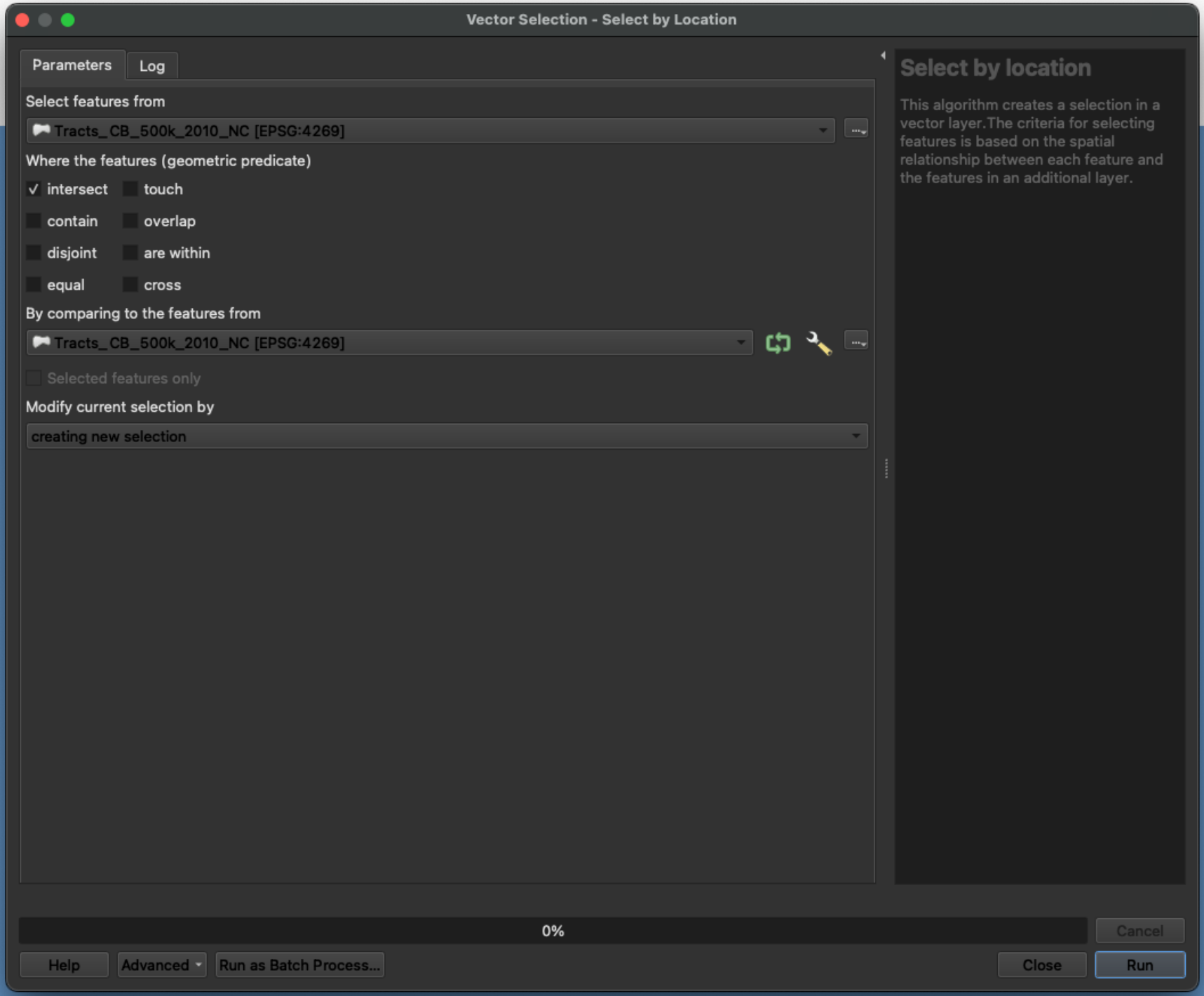

Tracts_CB_500k_2010_NC — Select by Expression
Expression
Function Editor
Search...
Show Help
Conditionals
Conversions
Date and Time
Fields and Values
feature
geometry
id
NULL
123 fid
abc GEO_ID
abc STATE
abc COUNTY
abc TRACT
abc NAME
abc LSAD
1.2 CENSUSAREA
abc COUNTYFP
abc STATEFP
Files and Paths
Fuzzy Matching
General
Geometry
Map Layers
Maps
Math
Operators
= + - / * ^ || ( ) '\n'
Feature 1
Preview:
Help
Select Features
Close

Selecting Features

- Select features by *location*
  - Two approaches to spatial query
    - Select a subset of entities through graphical interaction pointer
      - Box or lasso selection (user specified)
    - Select features in one data set based on distance or topological relationships with features in another data set

Select by Location

- Query based selection using spatial or topological relationships between layers
  - Topological concepts include
    - Containment, adjacency, proximity, intersection, crossing, shared features

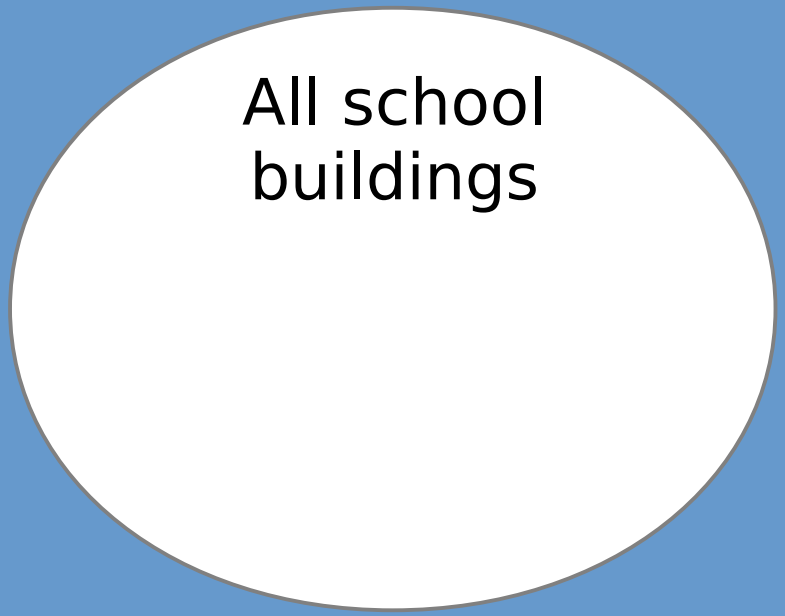

Vector Selection - Select by Location
Parameters
Log
Select features from
Tracts_CB_500k_2010_NC [EPSG:4269]
Where the features (geometric predicate)
intersect
touch
contain
overlap
disjoint
are within
equal
cross
By comparing to the features from
Tracts_CB_500k_2010_NC [EPSG:4269]
Selected features only
Modify current selection by
creating new selection
Select by location
This algorithm creates a selection in a vector layer.The criteria for selecting features is based on the spatial relationship between each feature and the features in an additional layer.
0%
Cancel
Help
Advanced
Run as Batch Process...
Close
Run

Combining Queries

- Important: by combining sequences of queries, including spatial and non-spatial queries, many higher level analytical operations can be performed
  - Think about:
    - What combination of operations would be needed to identify school buildings with predominant minority enrollment that are within 500 meters of a railroad line that carries toxic substances?

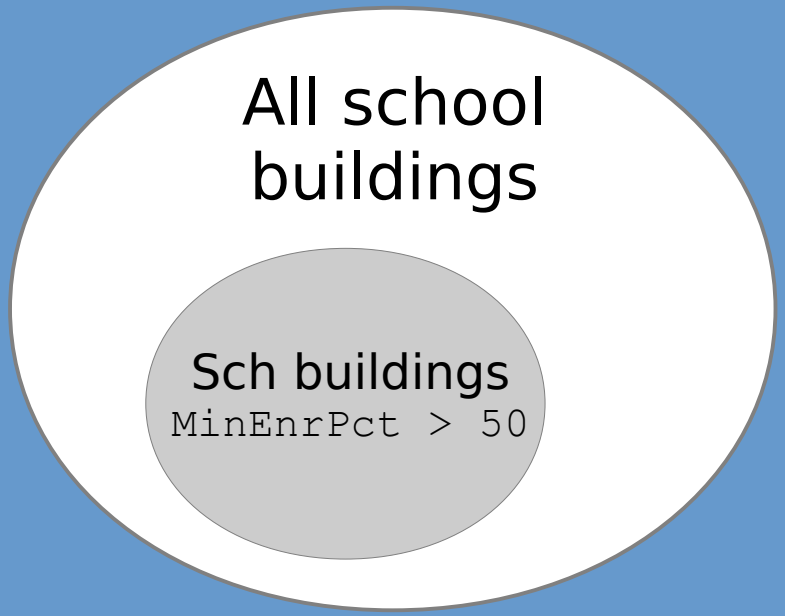
Combining Queries

- Identify *school buildings* with predominant minority enrollment that are within 500 meters of a railroad line that carries toxic substances?



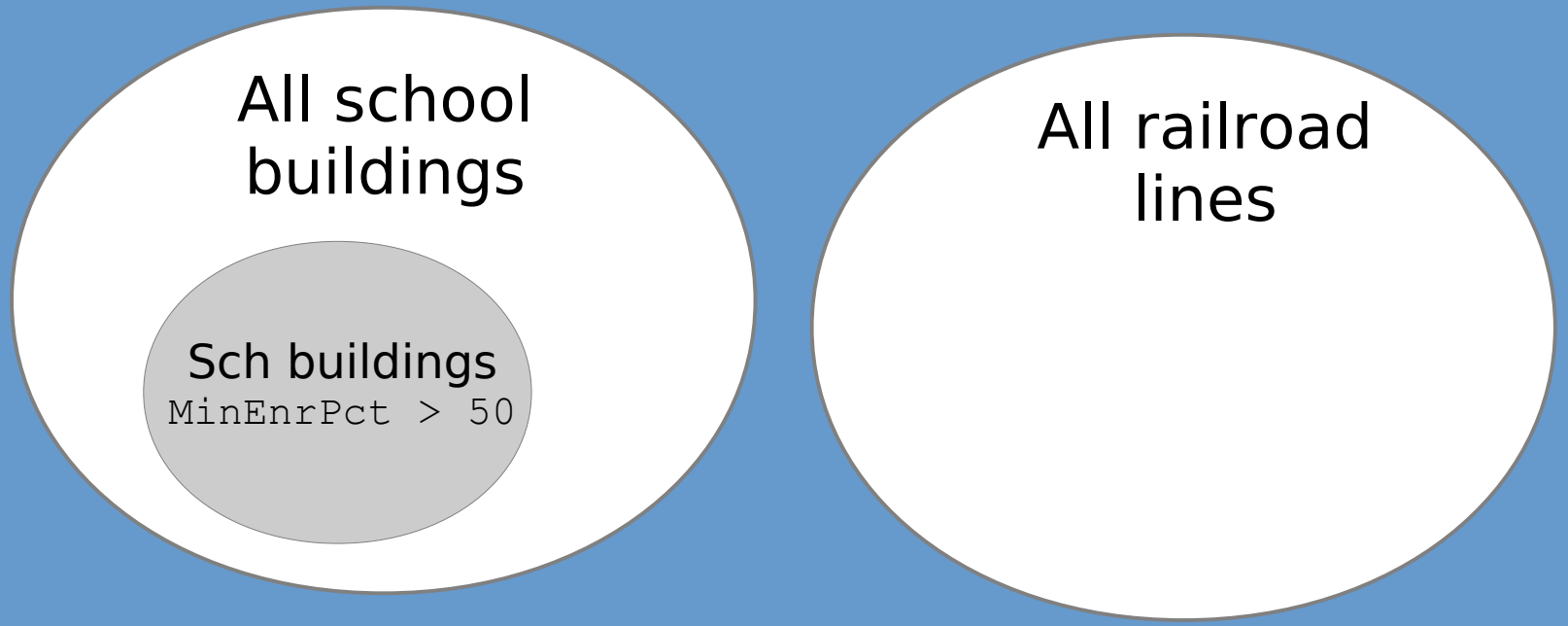
Combining Queries

- Identify school buildings *with predominant minority enrollment* that are within 500 meters of a railroad line that carries toxic substances?



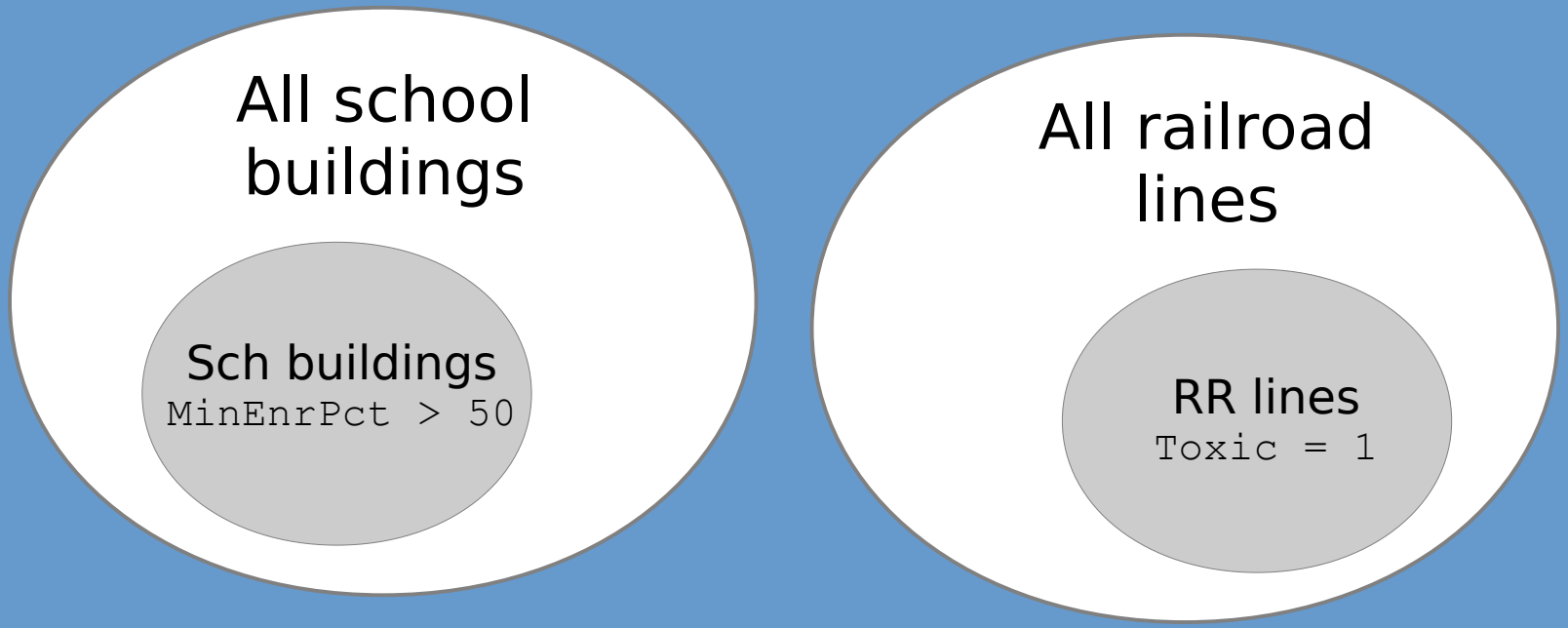
Combining Queries

- Identify school buildings with predominant minority enrollment that are within 500 meters of a *railroad line* that carries toxic substances?



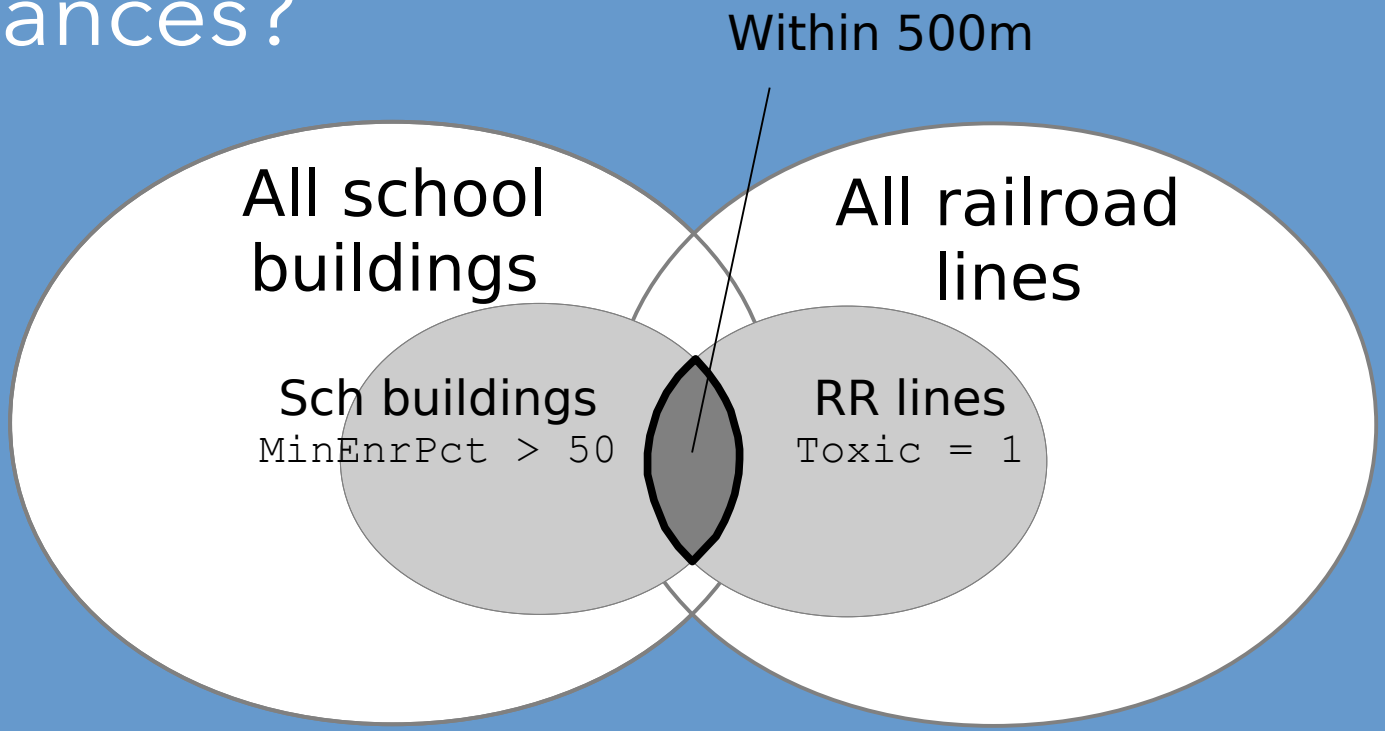
Combining Queries

- Identify school buildings with predominant minority enrollment that are within 500 meters of a railroad line that *carries toxic substances*?



Combining Queries

- Identify school buildings with predominant minority enrollment that are *within 500 meters* of a railroad line that carries toxic substances?



Data Selection In QGIS

- Selecting observations in a layer does not change the data in any way
  - Simply selects the observation(s)
- To create a new layer of selected features...
  - Select
  - Export data

Calculate Attributes

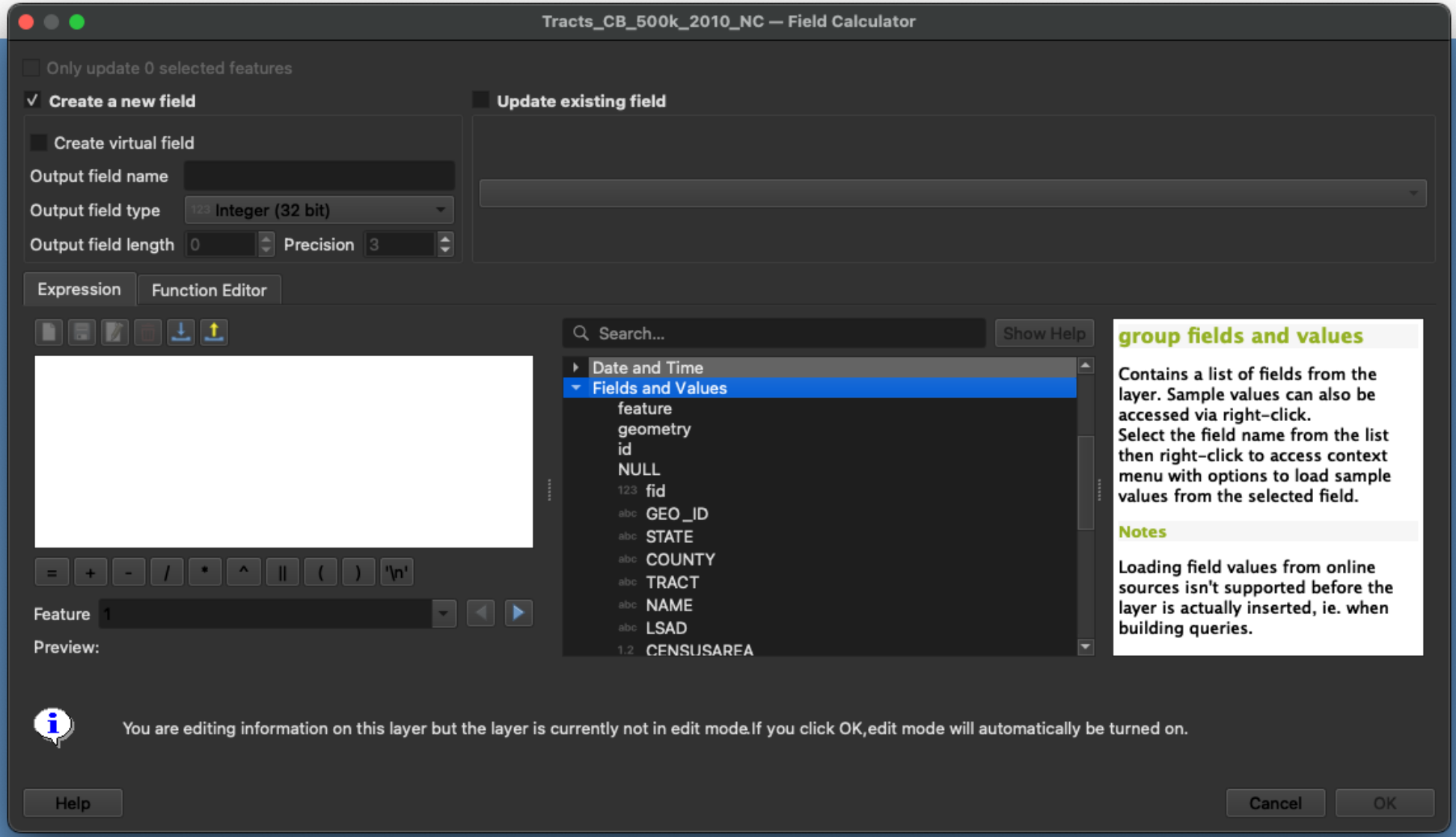
- Calculating attribute values allows the user to modify or combine existing attributes
  - Calculations can be applied to all features or selected subsets of features
    - To calculate new attribute values you can:
      - Define and create a new field in the table
      - Select the subset of records for which you want to calculate the new field
      - Enter attribute values or calculate the value of the new field as a function of existing attributes

Calculate Values

- General options:
  - Equivalence
    - e.g., `newfield = existfield`
  - Arithmetic expressions (+, -, *, /)
    - e.g., `newfield = existfield - 100`
  - Arithmetic functions (e.g., ln, cos, abs)
    - e.g., `newfield = abs(existfield)`

Calculate Values

- Multiple fields and combinations of the general options can be used
  - Allows for more complex calculations
    - e.g., Species density
      - `Density = Count / (Area * 100)`
    - e.g., Percent population change
      - `PopChPct = (PopNew - PopOld) / PopOld`


Tracts_CB_500k_2010_NC — Field Calculator
Only update 0 selected features
Create a new field
Update existing field
Create virtual field
Output field name
Output field type
Integer (32 bit)
Output field length
0
Precision
3
Expression
Function Editor
Search...
Show Help
Date and Time
Fields and Values
feature
geometry
id
NULL
fid
GEO_ID
STATE
COUNTY
TRACT
NAME
LSAD
CENSUSAREA
group fields and values
Contains a list of fields from the layer. Sample values can also be accessed via right-click.
Select the field name from the list then right-click to access context menu with options to load sample values from the selected field.
Notes
Loading field values from online sources isn't supported before the layer is actually inserted, ie. when building queries.
= + - / * ^ || ( ) '\n'
Feature
1
Preview:
You are editing information on this layer but the layer is currently not in edit mode.If you click OK,edit mode will automatically be turned on.
Help
Cancel
OK

Field Statistical Summary

- Display the *statistical* characteristics of a given field
  - Count - how many observations
  - Maximum - highest value
  - Minimum - lowest value
  - Sum - total value across all records
  - Mean - average value
  - Standard Deviation - square-root of variance
  - Nulls - number of null values

Field Summary

- Creates a new summary table
  - Summary statistics based on a user-chosen field in the attribute table
  - Mathematical operations
    - e.g., count, average, minimum, and maximum values
  - Summary table can then be joined back to spatial features via table join

Field Summary

- Example
  - Spatial join disease cases to counties
    - Output table of new layer contains the county that each disease case is located in (field called COUNTY)
  - In QGIS, Statistics by categories

Keywords

- Query
  - Simple
  - Compound
- Boolean operations
  - AND, OR, NOT
- Select features
- Attribute queries
  - Select by attribute
- Spatial queries
  - Select by spatial location or relationship
- Calculate attributes
  - Field calculator
- Field statistics
- Field summary