

Building a Spatial Database in PostgreSQL

David Blasby
Refractions Research

dblasby@refractions.net
<http://postgis.refractions.net>



Introduction

- *PostGIS* is a spatial extension for PostgreSQL
- *PostGIS* aims to be an “OpenGIS Simple Features for SQL” compliant spatial database
- I am the principal developer

Topics of Discussion

- Spatial data and spatial databases
- Adding spatial extensions to PostgreSQL
- OpenGIS and standards

Why PostGIS?

- There aren't any good open source *spatial* databases available
 - commercial ones are very expensive
- Aren't any open source spatial functions
 - extremely difficult to properly code
 - building block for any spatial project
- Enable information to be organized, visualized, and analyzed like never before

What is a Spatial Database?

Database that:

- Stores spatial objects
- Manipulates spatial objects just like other objects in the database

What is Spatial data?

- Data which describes either location or shape

e.g. House or Fire Hydrant location

Roads, Rivers, Pipelines, Power lines

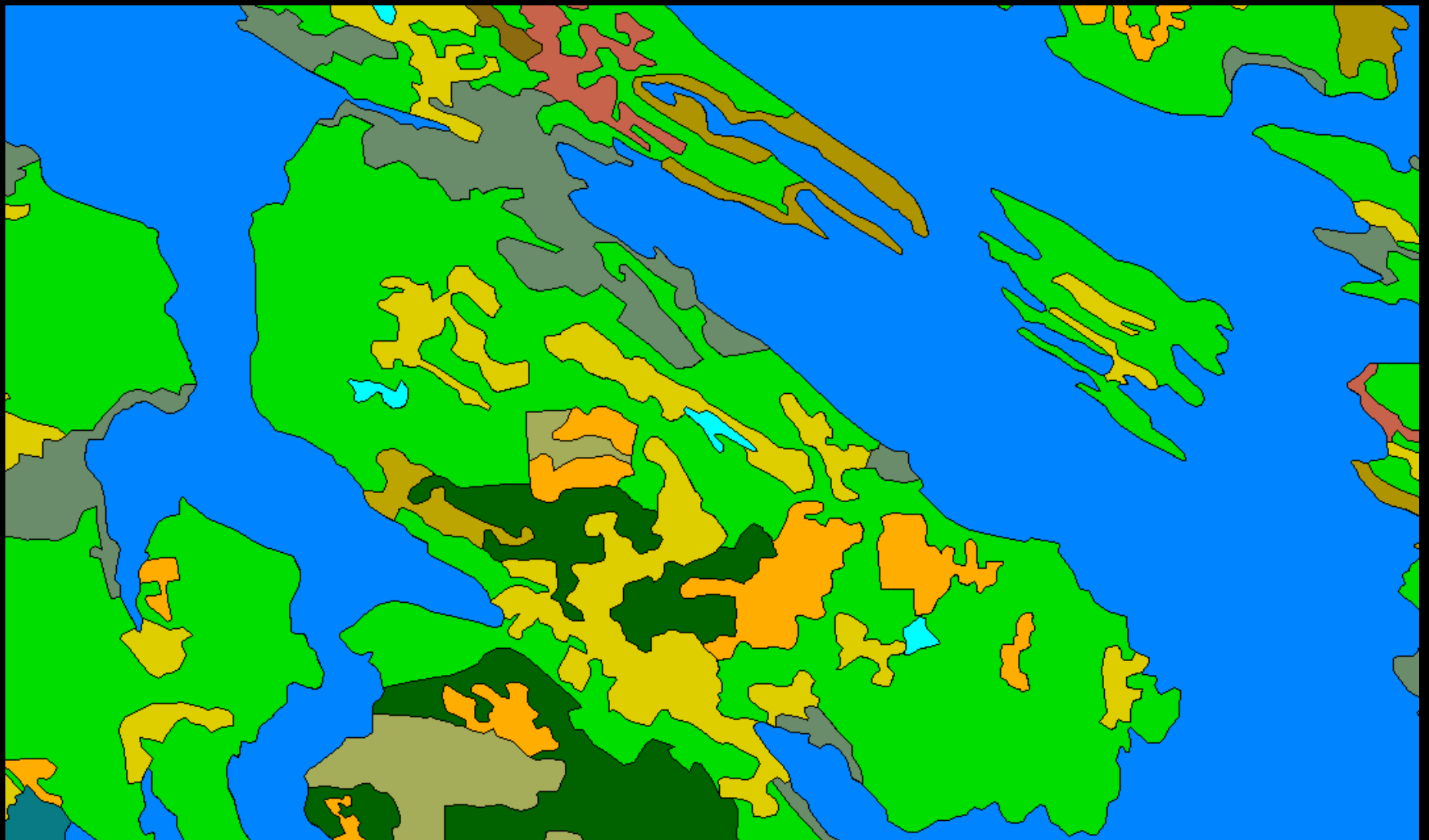
Forests, Parks, Municipalities, Lakes

What is Spatial data?

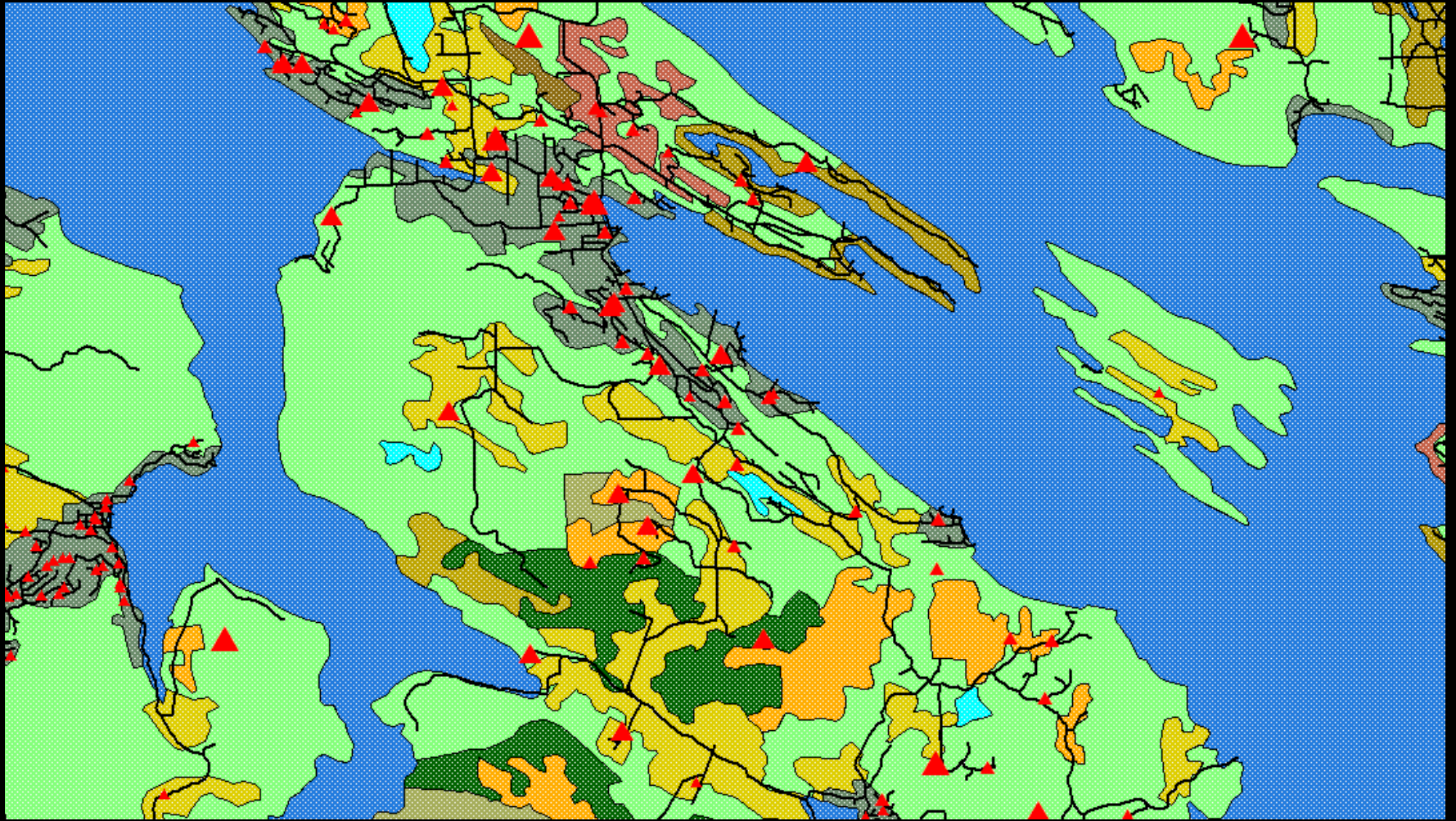
- In the abstract, reductionist view of the computer, these entities are represented as Points, Lines, and Polygons.



Roads are represented as Lines
Mail Boxes are represented as Points



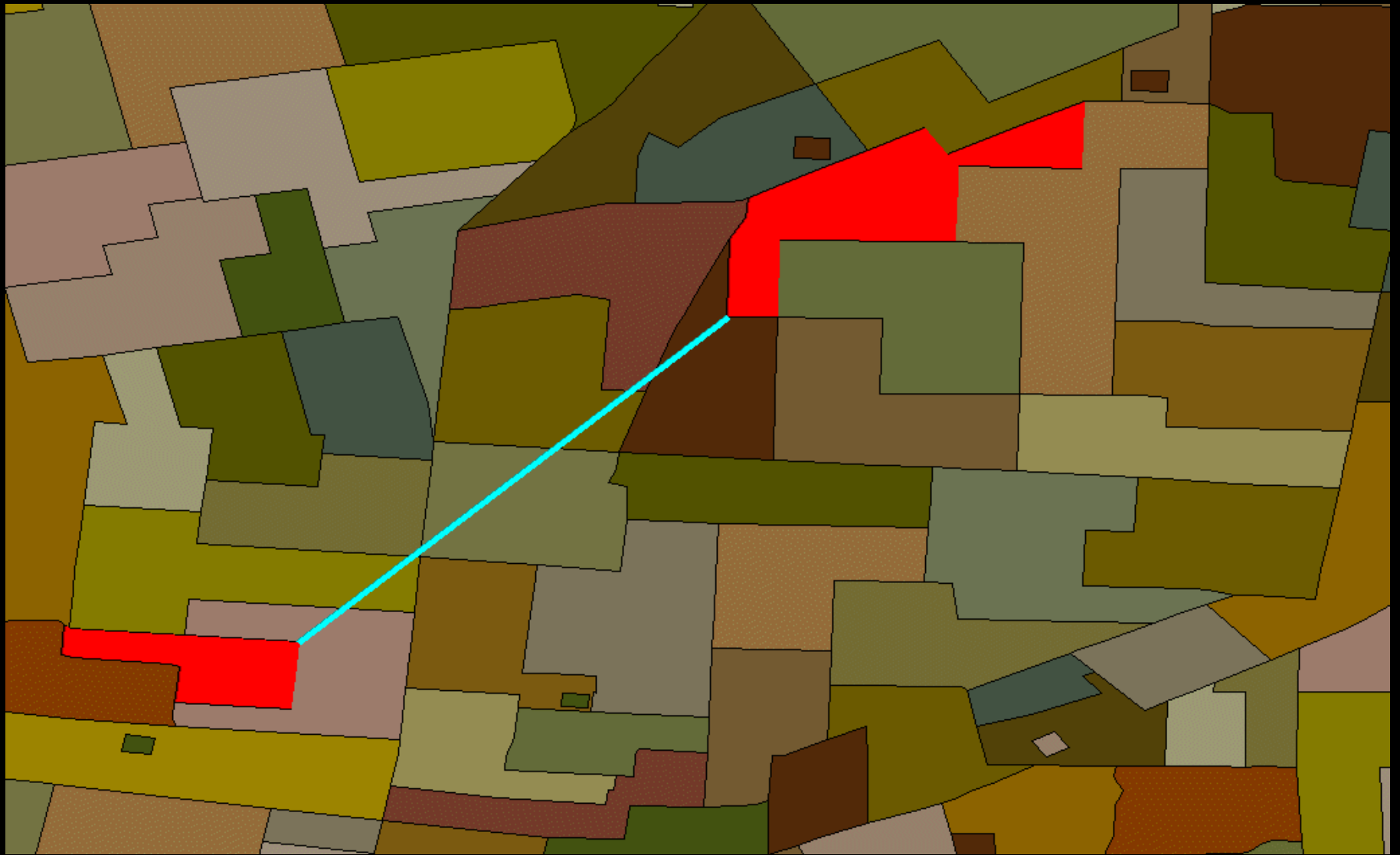
Land Use Classifications are
represented as Polygons



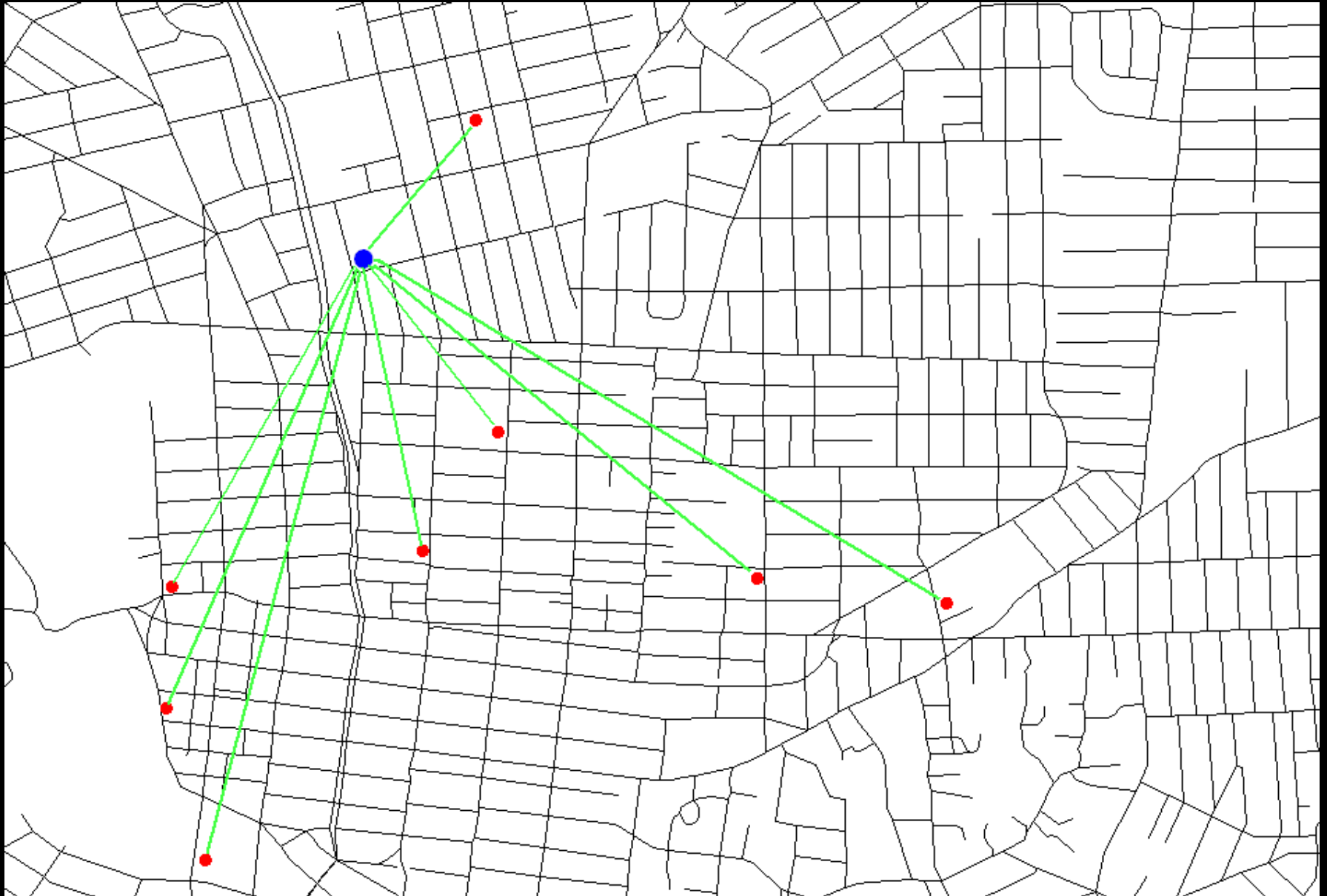
Combination of all the previous data

Spatial Relationships

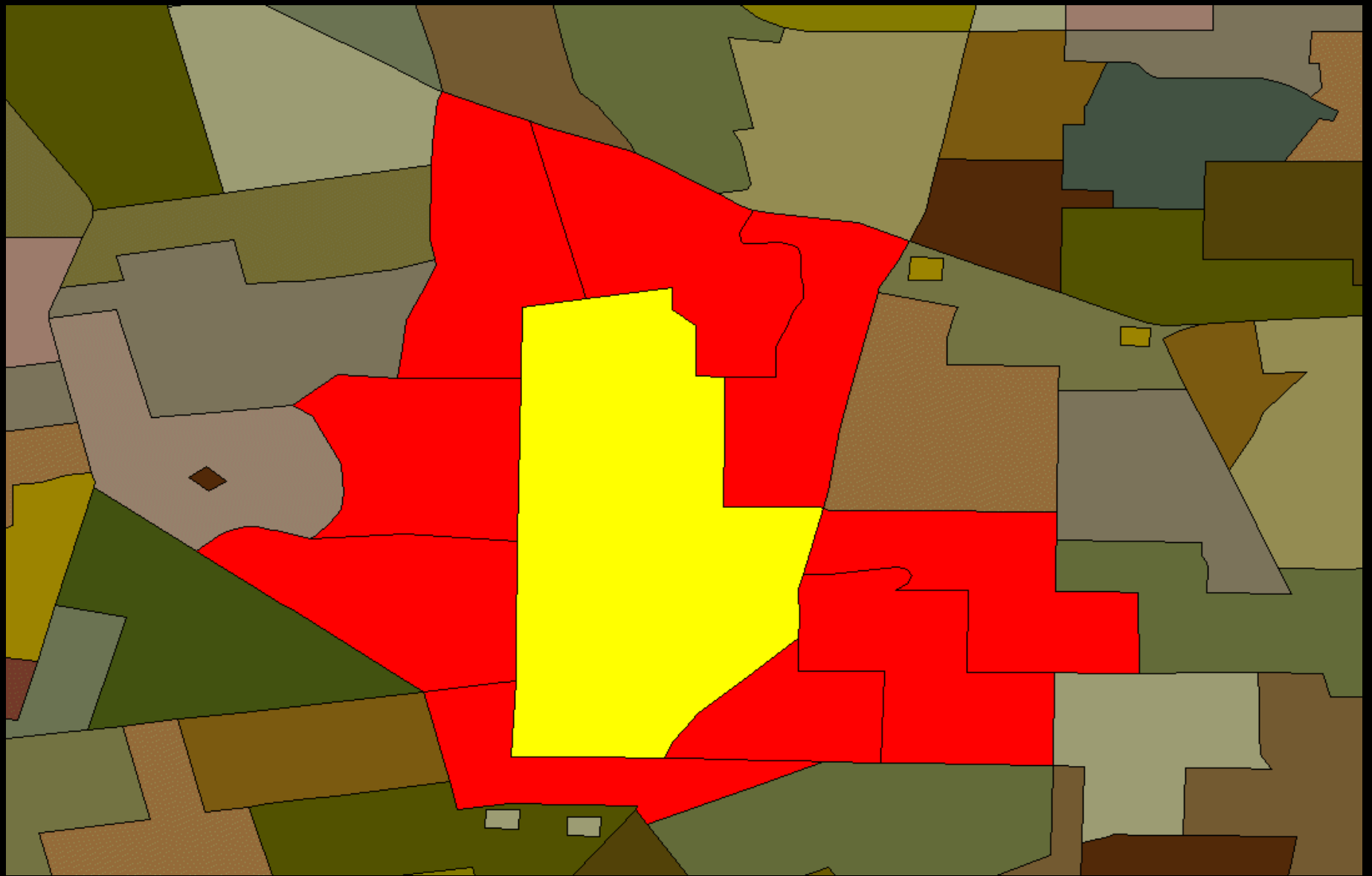
- Not just interested in location, also interested in “Relationships” between objects that are very hard to model outside the spatial domain.
- The most common relationships are
 - Proximity : distance
 - Adjacency : “touching” and “connectivity”
 - Containment : inside/overlapping



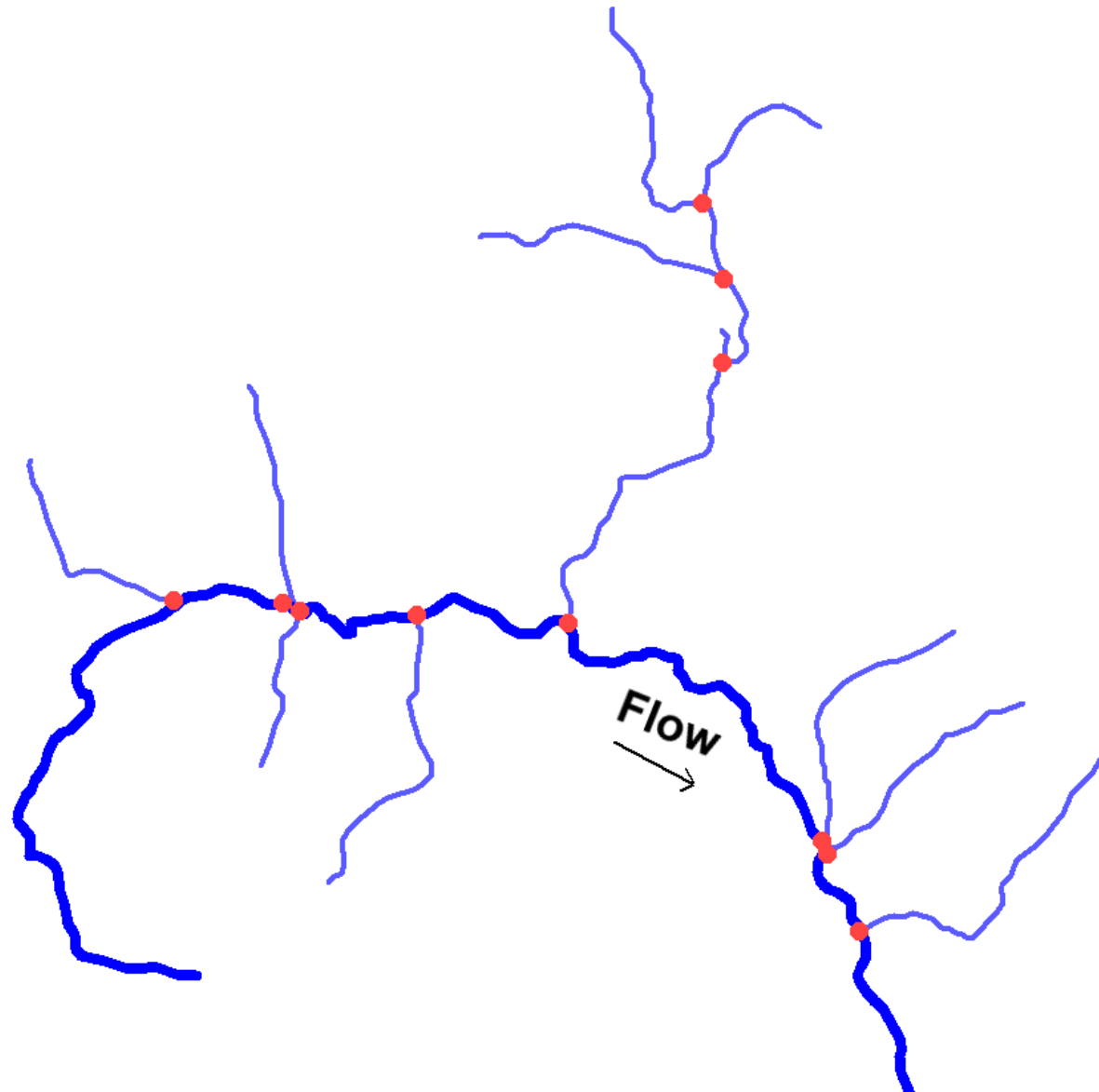
Distance between a toxic waste dump and a piece of property you were considering buying.



Distance to various pubs



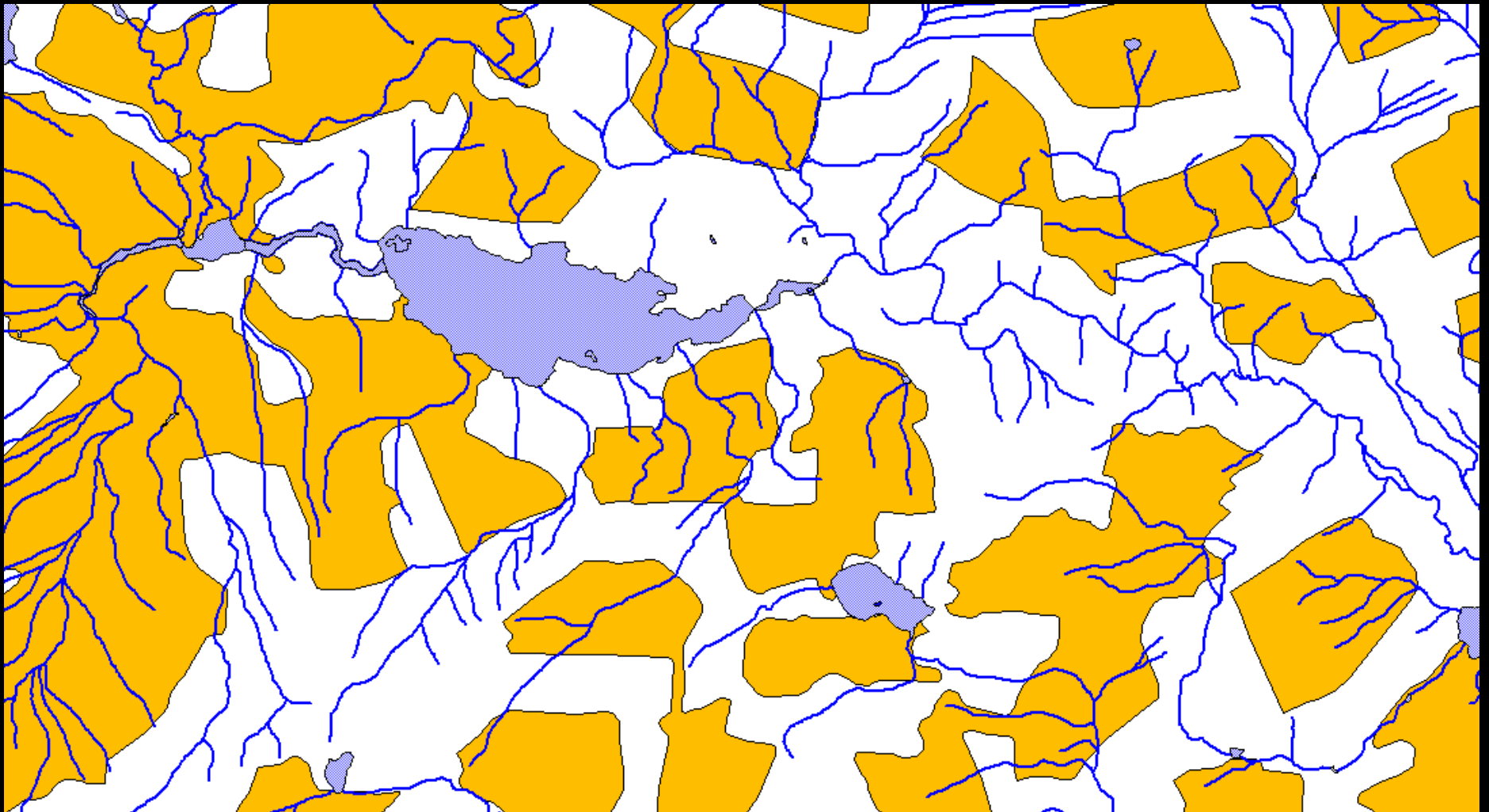
Adjacency: All the lots which share an edge



Connectivity: Tributary relationships in river networks



Containment: Rivers inside watersheds and land (islands) inside lakes



Stream side logging - adjacency and
containment.

Most Organizations have Spatial Data

- Geocodable addresses
- Customer location
- Store locations
- Transportation tracking
- Statistical/Demographic
- Cartography
- Epidemiology
- Crime patterns
- Weather Information
- Land holdings
- Natural resources
- City Planning
- Environmental planning
- Information Visualization
- Hazard detection

Why put spatial data in a RDBMS?

- Spatial data is usually related to other types of data. Allows one to encode more complex spatial relationships.
- Fire Hydrant: number of uses, service area, last maintenance date.
- River: flow, temperature, fish presence, chemical concentrations
- Forested Area: monetary value, types of trees, ownership

Historically?

- In early GIS implementations, spatial data and related attribute information were stored separately. The attribute information was in a database (or flat file), while the spatial information was in a separate, proprietary, GIS file structure.

For example, municipalities often would store property line information in a GIS file and ownership information in a database.

- Spatial databases were born when people started to treat spatial information as first class database objects.

Advantages of Spatial Databases

Able to treat your spatial data like anything else in the DB

- transactions
- backups
- integrity checks
- less data redundancy
- fundamental organization and operations handled by the DB
- multi-user support
- security/access control
- locking

Advantages of Spatial Databases

Offset complicated tasks to the DB server

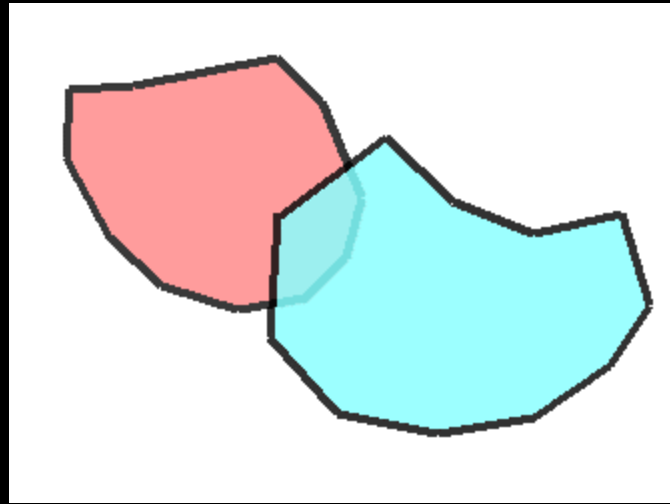
- organization and indexing done for you
- do not have to re-implement operators
- do not have to re-implement functions

Significantly lowers the development time of client applications

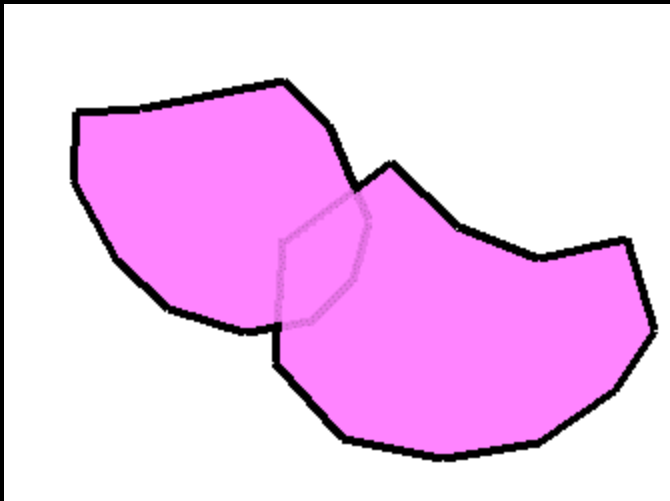
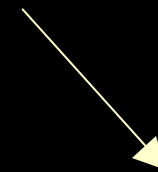
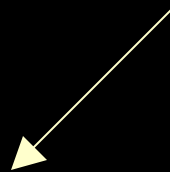
Advantages of Spatial Databases

Spatial querying using SQL

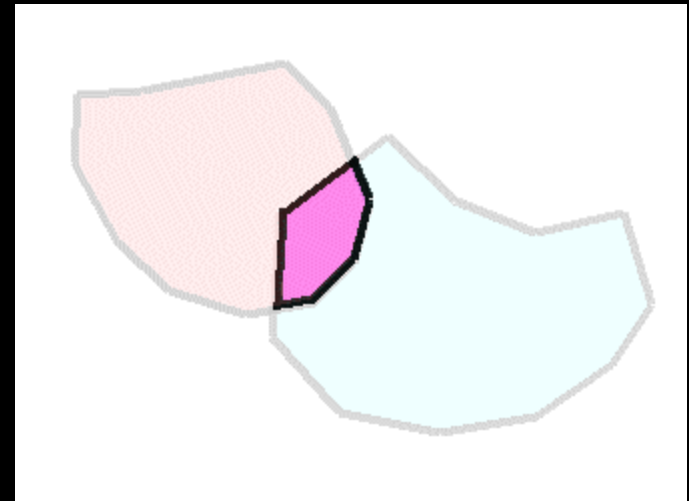
- use simple SQL expressions to determine spatial *relationships*
 - distance
 - adjacency
 - containment
- use simple SQL expressions to perform spatial *operations*
 - area
 - length
 - intersection
 - union
 - buffer



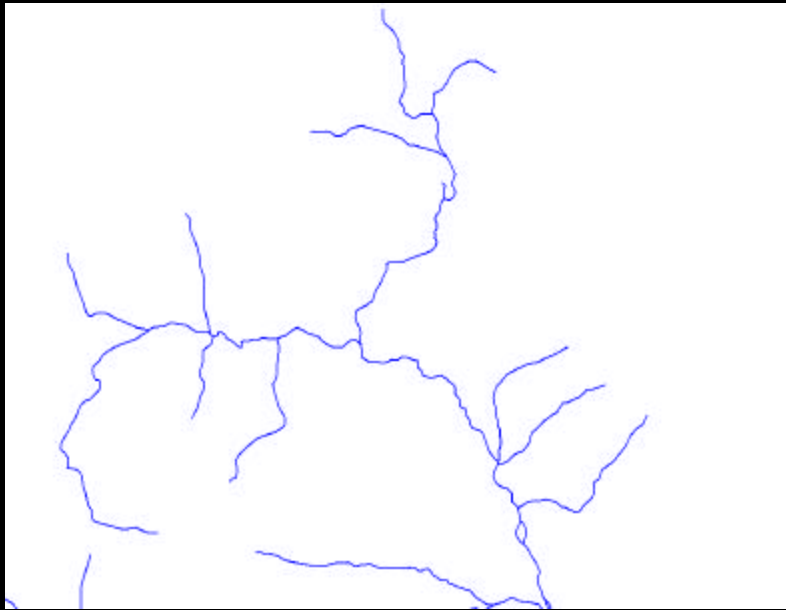
Original Polygons



Union

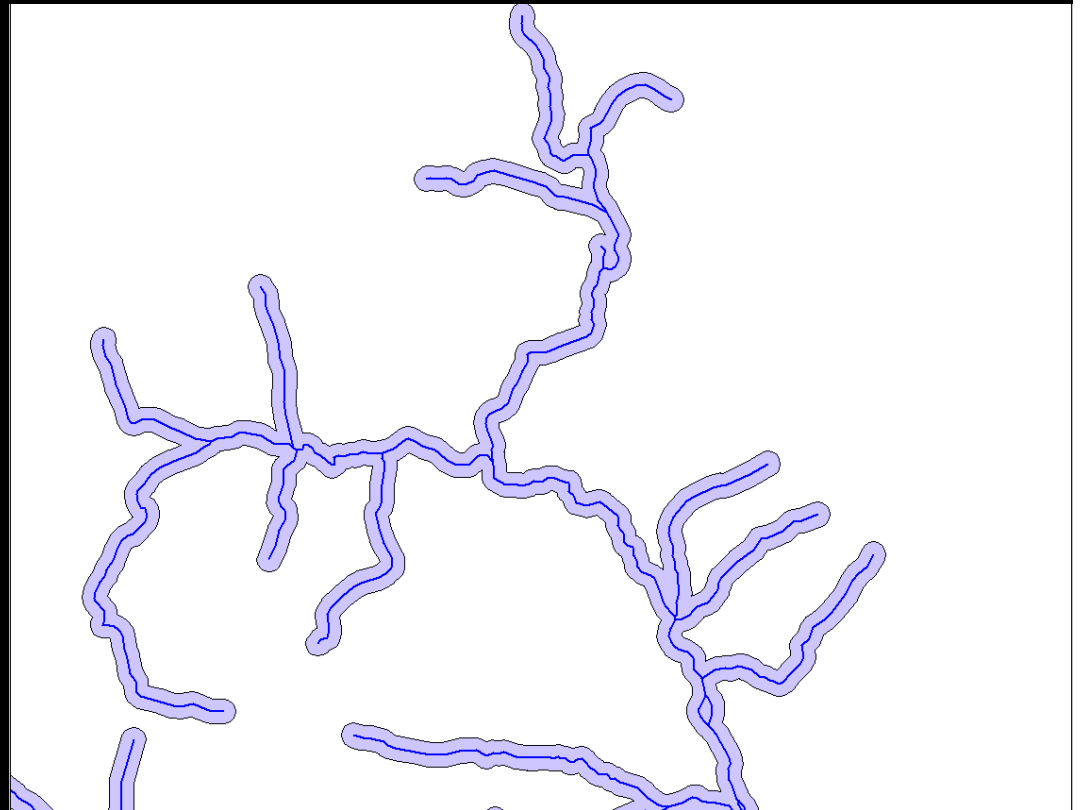


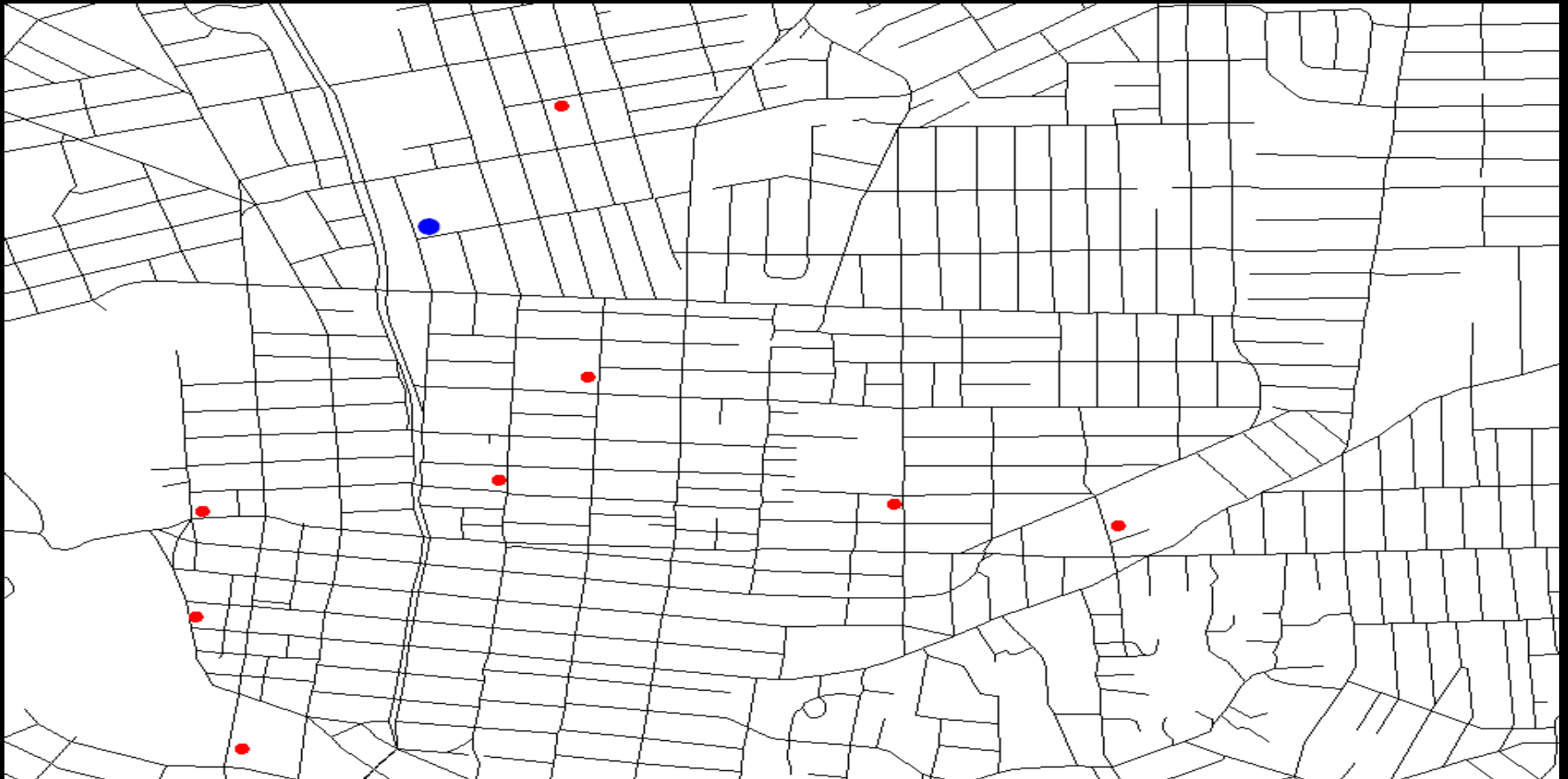
Intersection



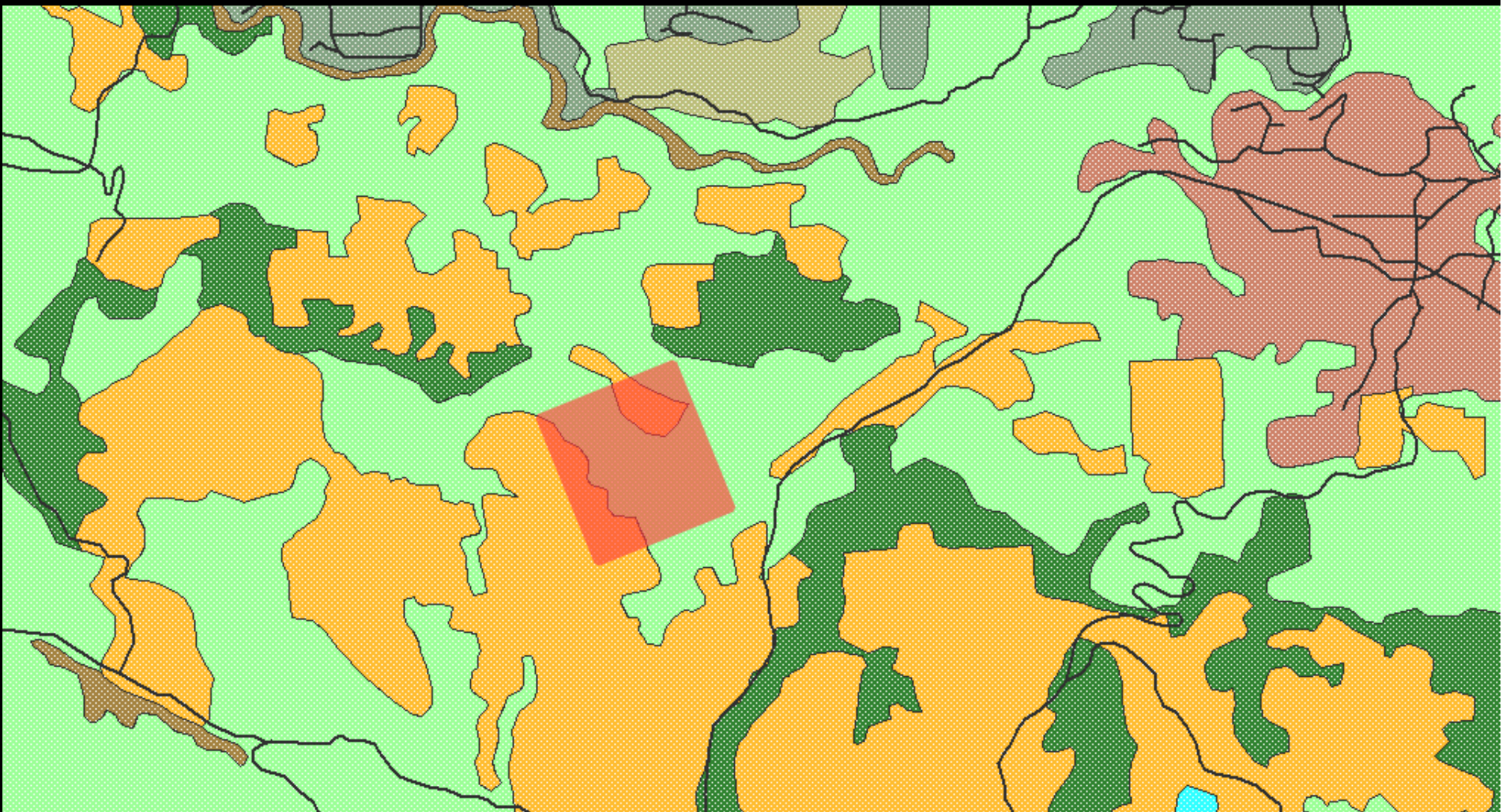
Original river network

Buffered rivers



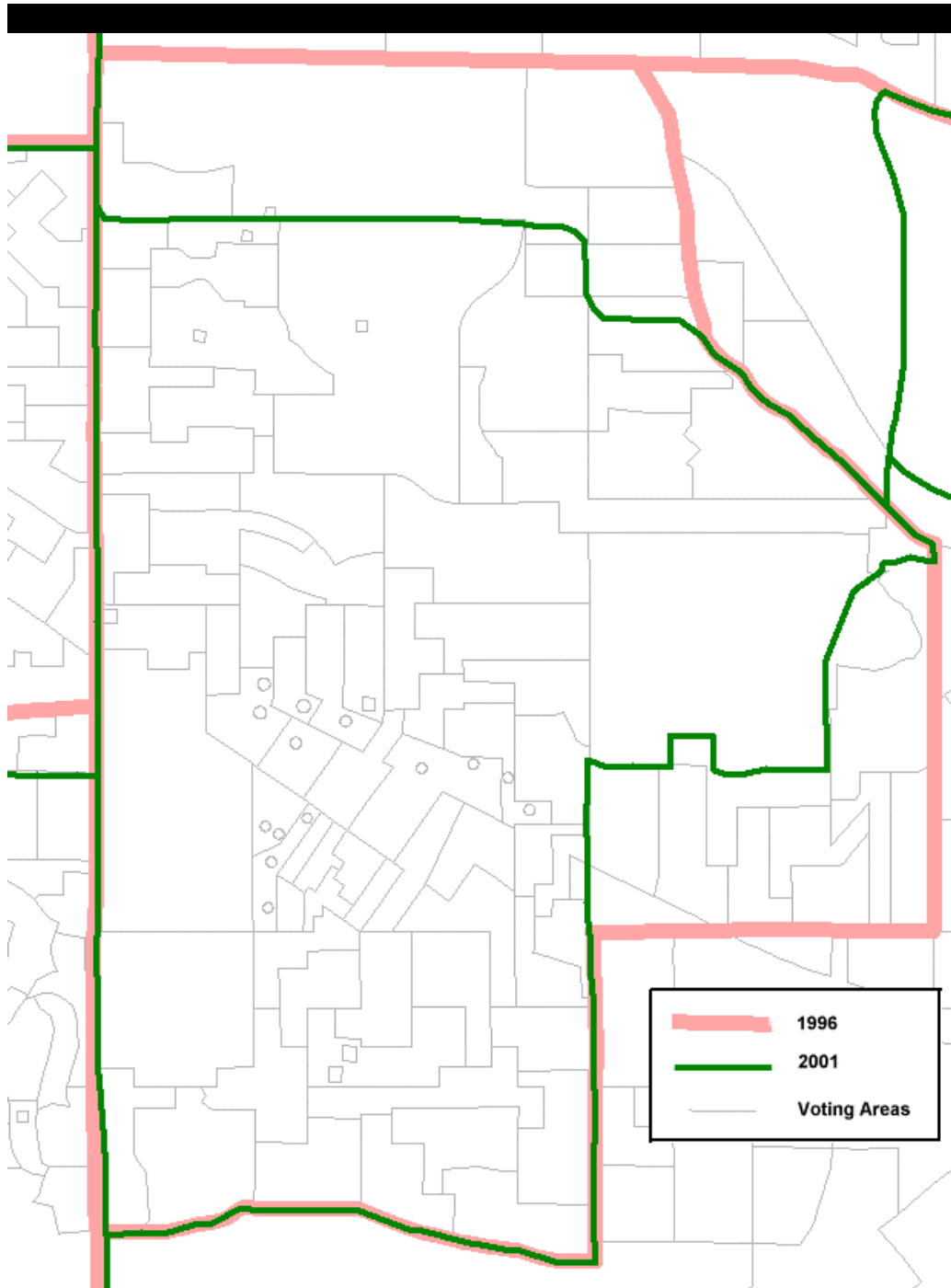


```
... WHERE distance(<me>,pub_loc) < 1000  
SELECT distance(<me>,pub_loc)*$0.01 + beer_cost ...  
... WHERE touches(pub_loc, street)  
... WHERE inside(pub_loc,city_area) and city_name = ...
```



Simple value of the proposed lot

$$\begin{aligned} & \text{Area}(\text{<my lot>}) * \text{<price per acre>} \\ & + \text{area}(\text{intersect}(\text{<my lot>}, \text{<forested area>})) * \text{<wood value per acre>} \\ & - \text{distance}(\text{<my lot>}, \text{<power lines>}) * \text{<cost of power line laying>} \end{aligned}$$



New Electoral Districts

- Changes in areas between 1996 and 2001 election.
- Want to predict voting in 2001 by looking at voting in 1996.
- Intersect the 2001 district polygon with the voting areas polygons.
 - Outside will have zero area
 - Inside will have 100% area
 - On the border will have partial area
- Multiply the % area by 1996 actual voting and sum
- Result is a simple prediction of 2001 voting

More advanced: also use demographic data.

Disadvantages of Spatial Databases

- Cost to implement can be high
- Some inflexibility
- Incompatibilities with some GIS software
- Slower than local, specialized data structures
- User/managerial inexperience and caution

Spatial Database Offerings

- ESRI ArcSDE (on top of several different DBs)
- Oracle Spatial
- IBM DB2 Spatial Extender
- Informix Spatial DataBlade
- MS SQL Server (with ESRI SDE)
- Geomedia on MS Access
- PostGIS / PostgreSQL

The OpenGIS Consortium

From the OpenGIS Consortium guide

“Much geospatial data is available on the web and in off-line archives, but it is complex, heterogeneous, and incompatible. Users must possess considerable expertise and special geographic information system (GIS) software to overlay or otherwise combine different map layers of the same geographic region. Data conversion is cumbersome and time-consuming, and the results are often unsatisfactory. Common interfaces are the only way to enable overlays and combinations of complex and essentially different kinds of geographic information to happen automatically over the Internet, despite differences in the underlying GIS software systems. OGC brings together the key players and provides a formal structure for achieving consensus on the common interfaces.”

The OpenGIS Consortium

From the OpenGIS Consortium FAQ

“OpenGIS is defined as transparent access to heterogeneous geodata and geoprocessing resources in a networked environment. The goal of the OpenGIS Project is to provide a comprehensive suite of open interface specifications that enable developers to write inter-operating components that provide these capabilities.”

Who is involved in the OpenGIS Consortium?

- ESRI
- Oracle
- IBM and Informix
- Intergraph
- Bentley (Microstation)
- MapInfo
- MicroSoft
- AutoDesk

Important OpenGIS Publications

- Simple Features Specification
 - for OLE/COM
 - for COBRA
 - for SQL
- Web Map Server Specification (WMS)
- Web Feature Server Specification (WFS)

Why make PostgreSQL into an OpenGIS SFSQL Spatial DB?

- Why choose PostgreSQL?
 - Proven reliability and respect
 - No cost (open source)
 - Supports most of the SQL standard
 - Ability to add new data-types
 - TOAST - no limit on column size
 - GiST index / Index extensions
 - Easy to add custom functions

Why make PostgreSQL into an OpenGIS SFSQL Spatial DB?

- Why choose OpenGIS SFSQL?
 - Third Party reviewed roadmap
 - One of the only open, respected standards
 - Participation by the major GIS/DB organizations
 - Other spatial DB are at least partially compliant
 - Future interoperability/exchangeability with other DBs
 - SFSQL provides complex functionality required by Web Feature Server / Web Map Server
 - User familiarity

Implementing the OpenGIS specification.

- Understand the Specification
 - Much harder than it sounds
- Add a GEOMETRY data type
 - Point / Multipoint
 - Linestring / Multilinestring
 - Polygon / Multipolygon
 - GeometryCollection
- Add support functions (and types)

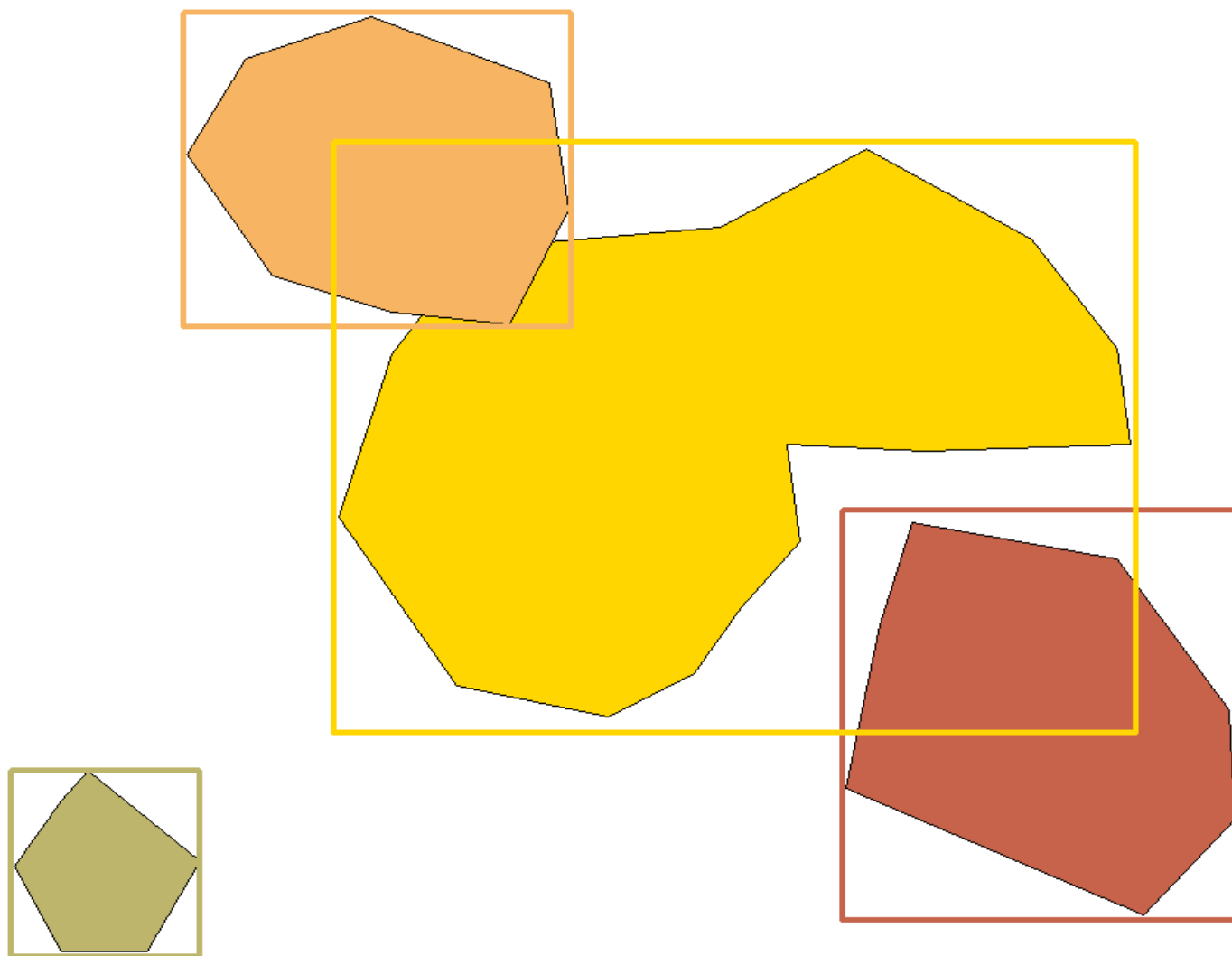
Spatial Indexing

Used the GiST (Generalized Search Tree) index

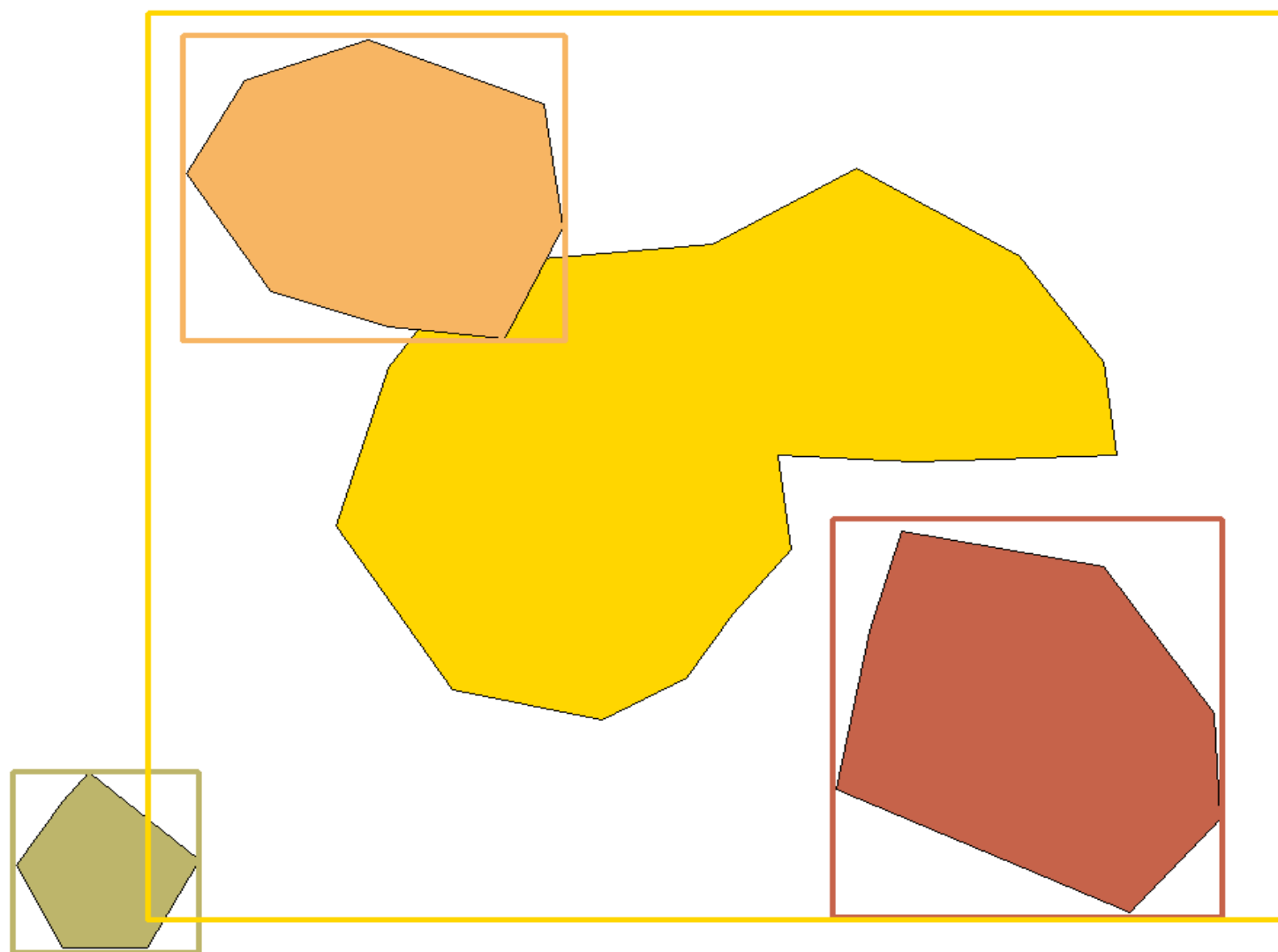
- Actively being developed
 - Teodor Sigaev and Oleg Bartunov
 - <http://www.sai.msu.su/~megera/postgres/gist/>
- Fast index creation
- Handles compression
 - use bounding box of the feature
- NULL safe
- Can implement an R-Tree using GiST

R-Tree Indexing

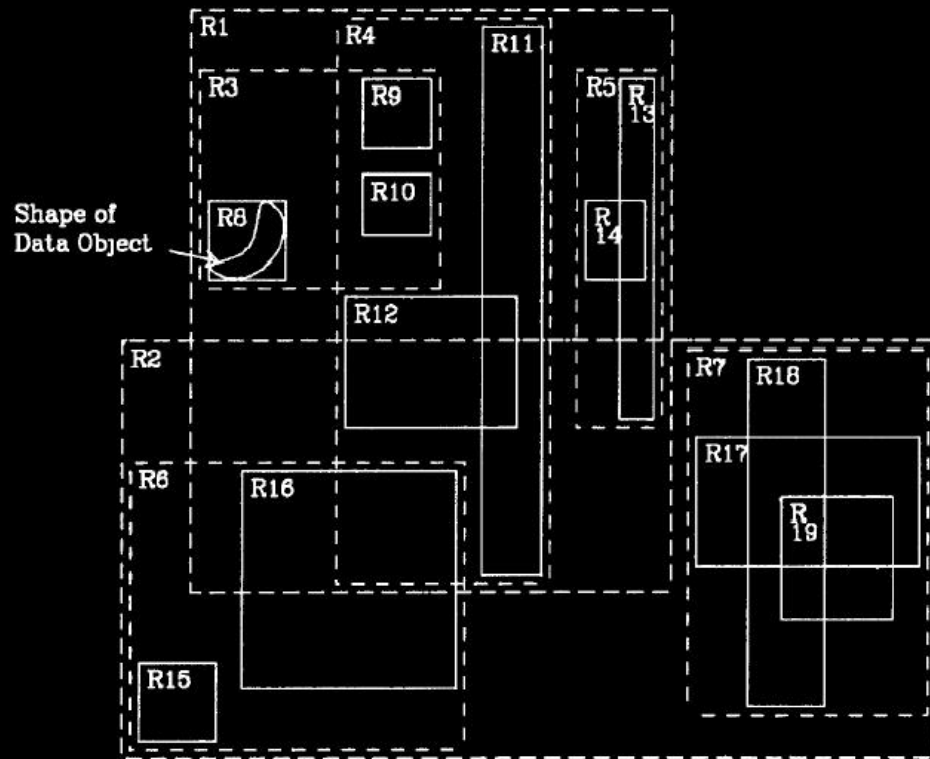
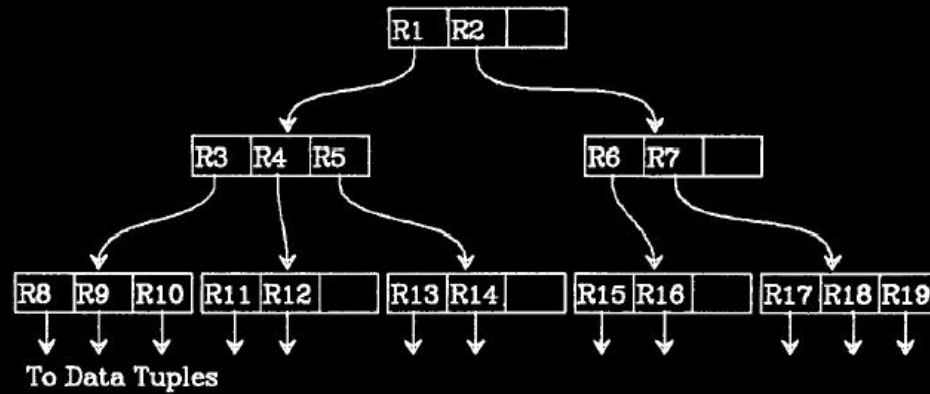
- Generalize all the geometries to their bounding box.
 - small to store
 - operations are simple
- Typical search is to find all the objects that overlap a box
- Result is an approximation
 - too many features are returned
- Used to solve overlap and distance problems



Overlap



Distance



Guttman A.: 'R-trees: A Dynamic Index Structure for Spatial Searching', Proc ACM SIGMOD Int. Conf. on Management of Data, 1984

SQL example

Create “pubs” table

```
create table pubs (name varchar,  
                  beer_price float4);
```

```
addgeometrycolumn('beer_db', 'pubs', 'location'  
                  , 2167, 'POINT', 3);
```



Insert data

```
insert into pubs values (  
    'Garricks Head',  
    4.50,  
    GeometryFromText(  
        'POINT (1196131 383324)'    ,2167)  
    );
```

Perform Query

```
select name, beer_price,  
       distance(location, GeometryFromText('POINT(1195722  
383854)',2167))  
       from pubs order by beer_price;
```

name	beer_price	distance
-----+-----+-----		
Fireside	4.25	1484.10275160491
The Forge	4.33	1533.06561109862
Rumours	4.46	2042.00094093097
Garricks Head	4.5	669.389105609889
Slap Happy	4.5	1882.31910168298
Old Bailys	4.55	1147.20900404641
Black Sheep	4.66	536.859935972633
Big Bad Daves	4.75	907.446543878884

Perform Query

```
select name, beer_price + 0.001 * distance(location,  
GeometryFromText('POINT(1195722 383854)',2167)) as net_price  
from pubs order by price;
```

name	net_price
-----+-----	
Garricks Head	5.16938910560989
Black Sheep	5.19685978338474
Big Bad Daves	5.65744654387888
Old Bailys	5.69720919478127
Fireside	5.73410275160491
The Forge	5.86306553480468
Slap Happy	6.38231910168298
Rumours	6.50200097907794



Client Software

What talks to PostGIS?

- Uses standard SQL so can connect to it from any client
- FME (Safe Software): GIS translation/processing
- Mapserver (<http://mapserver.gis.umn.edu>), an OpenGIS Web Map Server
- OGR (<http://gdal.velocet.ca/projects/opengis/>) - open source GIS reader/writer
- ESRI shapefile reader/writer
- In progress: ESRI ArcGIS connection, AutoCAD, Java Viewer, Web Feature Server

Open Standards in a Proprietary World

- Biggest obstacle is that most GIS companies have a closed/proprietary method for accessing and organizing spatial data
- ERSI's SDE (US\$10,000) is required to effectively connect its software to a spatial database.

“As explained above, ArcSDE is the gateway to the DBMS for ESRI's client applications. Without ArcSDE, customer sites are limited in what they can do with their spatial databases. ” - ESRI's ARC-SDE FAQ (www.esri.com)



Status

PostGIS 0.6 was released last week

- Implements all of the OGC specification except the “hard” spatial operations.
- Over 500 downloads and many people actually using it

The PostGIS development team is working with Vivid Solutions to include the Java Topology Suite (JTS). The JTS is an open source, rigorous, and robust implementation of OGC compliant spatial operations.

- Martin Davis
 - <http://www.vividsolutions.com/jts/jtshome.htm>

Conclusions

- PostGIS spatially enables PostgreSQL by adding spatial objects, functions, and indexing.
- PostGIS is free software (GPL)
- PostGIS follows the OpenGIS Simple Features for SQL
 - hope it will be certified next year
- PostGIS is an important component in open and free GIS.
- PostGIS is an important building block for all future open source spatial projects.

Questions

David Blasby
Refractions Research

dblasby@refractions.net
<http://postgis.refractions.net>

