

# Benchmark Testing of GIS Capabilities for Epidemiologic Studies of Breast Cancer on Long Island

Procedural considerations and selected  
functional tests for benchmarking a GIS

Frederick R. Broome  
Geography Division  
US Bureau of the Census  
1999.11.15

# The Long Island Breast Cancer GIS

While the Long Island Breast Cancer Study Project GIS is in the title, the principles and procedures described are from several federal GIS procurement activities and are not unique to the LIBCSP.

Indeed, the benchmarking technique include examples from other procurements.

Therefore, the techniques presented are applicable to selecting a GIS for use in many areas.

---

For detailed information about the LIBCSP GIS, contact:

Dr. Ellen Heineman [heinemae@epndce.nci.nih.gov](mailto:heinemae@epndce.nci.nih.gov)

# Public Law 103-43: The First Federally Legislated GIS

Enacted in 1993

Directed, among other things, that the study "**should include the use of a geographic system** to evaluate the current and past exposure of individuals, including direct monitoring and cumulative estimates of exposure to (1) contaminated drinking water; (2) sources of indoor and ambient air pollution, including emissions from aircraft; (3) electromagnetic fields; (4) pesticides and other toxic chemicals; (5) hazardous and municipal waste; and (6) such other factors as the director [of NCI] determines to be appropriate."

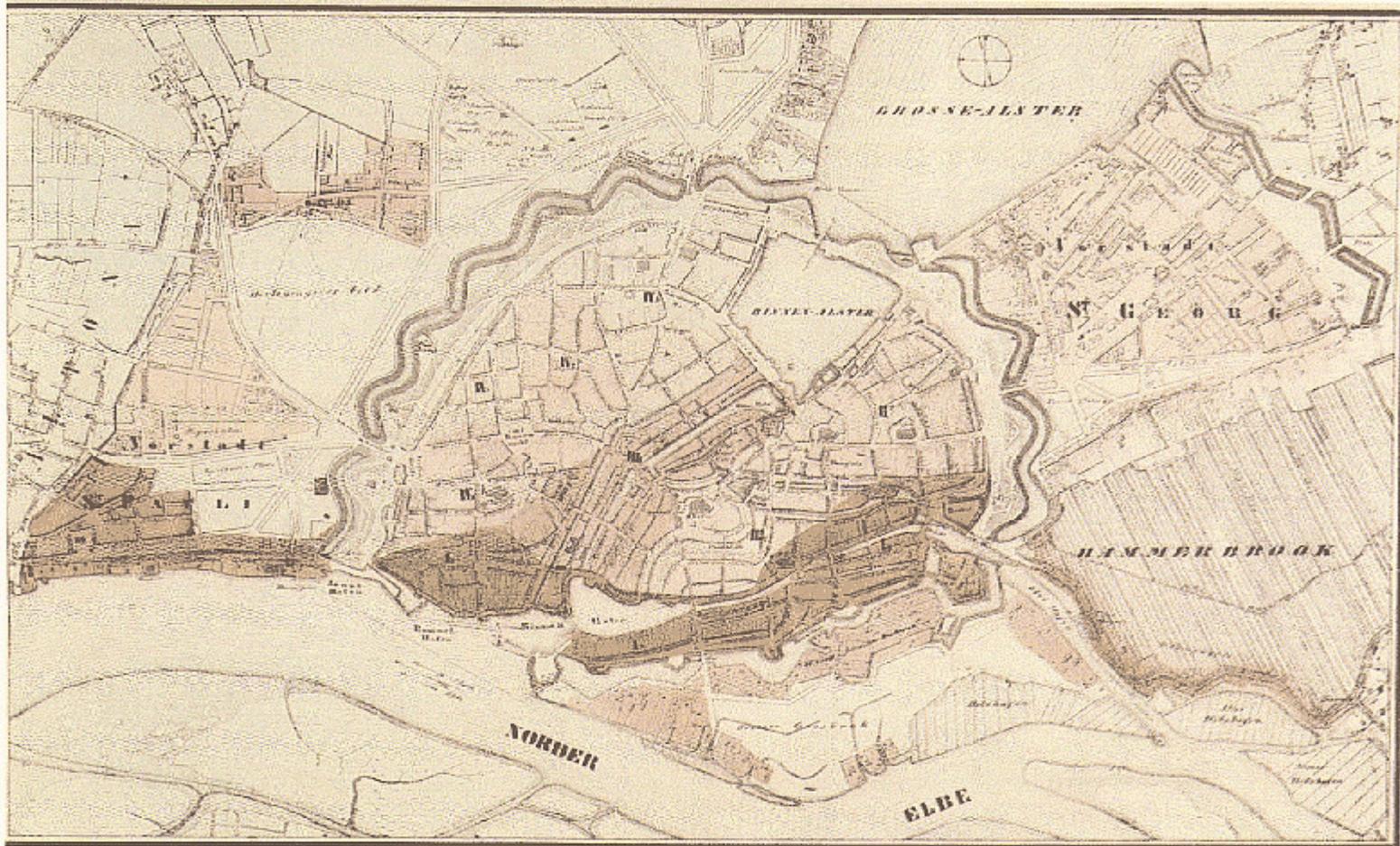
*Source: NIH Web Site*

# Use of GIS Techniques in Epidemiologic Studies is not New

Dr. Snow's  
London  
Deaths  
from  
Cholera  
Map, 1855



# Rothenburg's 1836 Map of Cholera



**Figure 83** Rothenburg's 1836 map accompanying his *Die Cholera Epidemie des Jahres 1832 in Hamburg*. Original 446 × 284 mm. Lithograph, hand colored. (Courtesy of the British Library.)

# GIS for Epidemiologic Studies Today

The question is not, “Can GIS technology be applied to epidemiologic studies today?”

The question is, “How does one choose the GIS?”

# Why Benchmark?

---

Product literature and demonstrations are designed to ‘put the best foot forward’ and to encourage sales.

Benchmarks are designed to test the product **against specific needs of the user**, that is, prove the product fulfills the user’s operational needs, and, if so, how and how efficiently.

# Benchmark Test Development Issues

---

- ❑ Identification and selection of project teams
- ❑ Selecting test data types and geographic level
- ❑ Determining appropriate tasks for evaluation
- ❑ Example datasets for the tasks
- ❑ Example benchmark tasks
- ❑ Importance of a trial run
- ❑ Benchmark test administration
- ❑ Evaluating performance

# Benchmark Test Development Issues

- Identification and selection of project teams
- Selecting test data types and geographic level
- Determining appropriate tasks for evaluation
- Example datasets for the tasks
- Example benchmark tasks
- Importance of a trial run
- Benchmark test administration
- Evaluating performance

# Project Teams

---

- ❑ System's goal identification team
- ❑ System's specification development team
- ❑ Benchmark test suite development team
- ❑ Benchmark testing and evaluation team

# Project Teams

## System's goal identification team



- Sets the goals and broad specifications for the GIS
- Consists of people familiar with the mission of the organization and how GIS will fit into it. If possible, includes members from area responsible for managing the GIS.
- Should have representatives or members from procurement and technical areas, both GIS and applications.

# Project Teams

## System specification development team



- Develops detailed specifications for the GIS
- Consists of technical experts and application experts.
- Should be headed by member of the goals identification team
- Coordinates with procurement staff to assure specifications are properly formulated and expressed as functional requirements

# Project Teams

## Benchmark test suite development team



- Identifies which functional specifications can be tested during a timed benchmark and develops the benchmark tasks
- Consists of GIS and applications area experts and is headed by a member of the goals identification team.
- Responsible for creation of specific task instructions and associated inputs, including test datasets, documentation, and reference materials

# Project Teams



## Benchmark testing and evaluation team

- Administers the actual benchmark test, evaluates and reports results
- Consists of representatives of technical and applications areas (usually members from the other teams)
- Headed by representative from the procurement area, or someone with procurement authority

# Benchmark Test Development Issues

- Identification and selection of project teams
- Selecting test data types and geographic level
- Determining appropriate tasks for evaluation
- Example datasets for the tasks
- Example benchmark tasks
- Importance of a trial run
- Benchmark test administration
- Evaluating performance

# Test Suite Data Types

*Representative of, but not necessarily actual, datasets to be used in the final system*

- Spatial (Geographic) Data - provides the geographic framework for all tasks
- Statistical Data - provides statistics related to the geographic framework
- Data Documentation - describes the datasets in sufficient detail that the vendor will not have to refer to external sources to load the data

# Benchmark Test Development Issues

- ❑ Identification and selection of project teams
- ❑ Selecting test data types and geographic level
- ☑ Determining appropriate tasks for evaluation
- ❑ Example datasets for the tasks
- ❑ Example benchmark tasks
- ❑ Importance of a trial run
- ❑ Benchmark test administration
- ❑ Evaluating performance

# Test Suite Geographic Resolution

*Establishes the finest level of geographic resolution to be represented by the datasets*

- May vary by task
- May include higher, more general levels

Examples:

1. Housing districts used for geocoding
2. Census block groups for statistical analysis and mapping

# Benchmark Test Development Issues

- ❑ Identification and selection of project teams
- ❑ Selecting test data types and geographic level
- ❑ Determining appropriate tasks for evaluation
- ☑ Example datasets for the tasks
- ❑ Example benchmark tasks
- ❑ Importance of a trial run
- ❑ Benchmark test administration
- ❑ Evaluating performance

# The Datasets (used in the benchmark)

---

1990 Population Statistics from the Census's Standard Tape Files

1980 Population Statistics from the Census's Standard Tape Files

1990 BG Centroids

1980 BG Centroids

TIGER/Line '95

TIGER/Line '90

EPA Air Quality Monitoring point data

Scanned Map Image of a USGS 7.5" quadrangle

An Address List of plausible address

Scanned Map Image for adjudication of new streets

Evaluator Defined Areas (defined geography for statistics aggregation)

Evaluator Defined Statistics with geocodes and coordinates

# Preparation of Datasets

---

- ❑ Data generation (prepared, extracted or proposer produced)
- ❑ Challenges encountered in building datasets

# Benchmark Test Development Issues

- ❑ Identification and selection of project teams
- ❑ Selecting test data types and geographic level
- ❑ Determining appropriate tasks for evaluation
- ❑ Example datasets for the tasks
- ☑ Example benchmark tasks
- ❑ Importance of a trial run
- ❑ Benchmark test administration
- ❑ Evaluating performance

# Exercises for Evaluation

---

Tasks should:

- ❑ Reflect work to be done in project
- ❑ Test use of GIS and other relevant software
- ❑ Test understanding of geographic concepts and cartographic skill

# Task: Dataset Loading



# Task: Internet Downloading

The screenshot shows a Netscape browser window titled "United States Environmental Protection Agency - Netscape". The address bar contains "http://search.epa.gov/s97is.vts". The main content area displays "Public Access Server Search Results" from the EPA. A green box indicates that the query "Landview III" matched 44 documents out of 229375. Below this, there are links for "Refine Search" and "New Search". Further down, there are links for "Hide Summaries", "Hide PDFs", and "About PDF Files". The search results are presented in a table with three columns: Rank, Title/Summary, and Format. The first result is "LandView III Mapping Tool Information" with a summary and a URL. The second result is "LandView® III Fact Sheet" with a summary and a URL.

United States Environmental Protection Agency - Netscape

File Edit View Go Window Help

Back Forward Reload Home Search Guide Print Security Stop

Bookmarks Netsite: http://search.epa.gov/s97is.vts

Internet Lookup New&Cool

**EPA** United States Environmental Protection Agency

## Public Access Server Search Results

The "Highlight Search Terms" link displays the document with your search terms highlighted in red.  
Click on the ">" to move to the next highlight or the "<" to move to the previous highlight.

Your query "Landview III" matched 44 documents out of 229375.  
Documents 1 through 10 are listed below in order of relevance. [\[Refine Search\]](#) [\[New Search\]](#)

[\[Hide Summaries\]](#) [\[Hide PDFs\]](#) [\[About PDF Files\]](#)

Rank	Title/Summary	Format
1.	<a href="#">LandView III Mapping Tool Information</a> LandView IIITM: A Tool for Community Brownfields Projects Factsheet [ HTML (15K)   PDF (58K) 4 Pages ] EPA Document Number: EPA 500-F-98-006 June 1998 LandView III General Fact Sheet CEPPPO Factsheet [ HTML (9K)   PDF (264K) 4 Pages ] EPA Document Number E URL: <a href="http://www.epa.gov/swerosps/bf/lvinfo.htm">http://www.epa.gov/swerosps/bf/lvinfo.htm</a> <a href="#">Highlight Search Terms</a>	
2.	<a href="#">LandView® III Fact Sheet</a> LandView® III is an innovative "Community Right-To-Know" software tool and the most recent in a series of electronic tools developed by EPA's Chemical Emergency Preparedness and Prevention Office (CEPPPO). LandView® III provides database extracts from the URL: <a href="http://www.epa.gov/swerosps/bf/html-doc/lvfactcpo.htm">http://www.epa.gov/swerosps/bf/html-doc/lvfactcpo.htm</a> <a href="#">Highlight Search Terms</a>	

Document Done

# Task: Address Matching

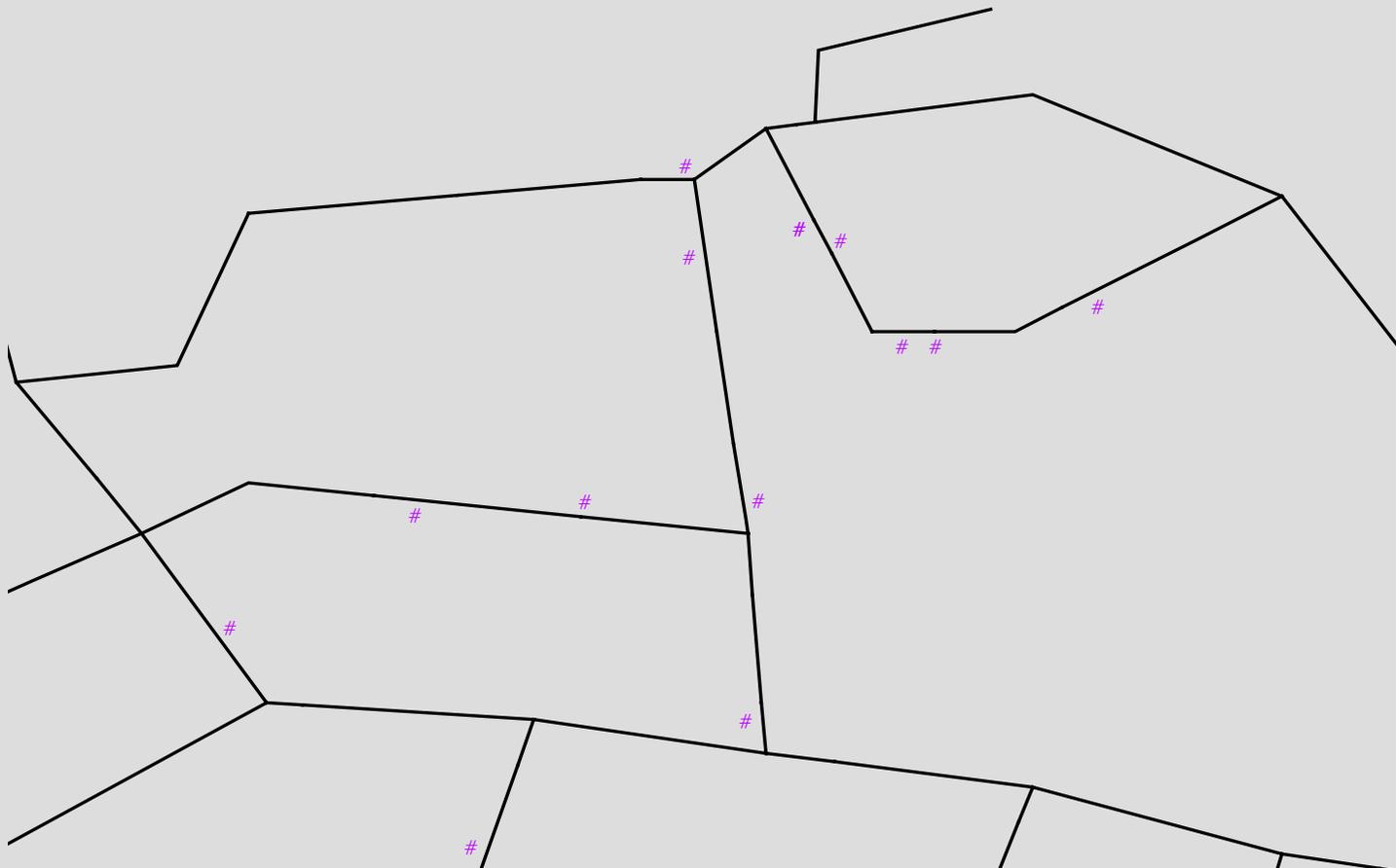
Portion of addresses in file to be used in address matching task.

```
AddressTypeID, StructureAddress, StreetID, Place, State, ZIP
Occupant 010001, 36, Pinecre Ct, Columbia, SC, 29204
Occupant 070002, 152, Chimney Ridge Dr, Columbia, SC, 29223
Occupant 060003, 1718, Shop Rd, Columbia, SC, 29201
Occupant 010004, 7028, Crossfield Rd, Columbia, SC, 29206
Occupant 010005, 5607, Colonial Dr, Columbia, SC, 29203
Occupant 030006, 1393, Whitaker Dr, Columbia, SC, 29206
Occupant 050007, 2085, Rolling Hills Rd, Columbia, SC, 29210
Occupant 090008, 9870, Caughman Rd, Columbia, SC, 29209
Occupant 040009, 180, Running Fox Rd, Columbia, SC, 29210
Occupant 050010, 1313, Cullum St, Columbia, SC, 29209
Occupant 020011, 1480, Old Hilton Rd, Columbia, SC, 29036
Occupant 050012, 366, Lawand Dr, Columbia, SC, 29210
Occupant 060013, 12, Myron Cir, Columbia, SC, 29209
Occupant 020014, 1188, Fresh Mill Rd, Columbia, SC, 29063
Occupant 060015, 1002, Timrod St, Columbia, SC, 29203
Occupant 010016, 7344, Spring Bank Rd, Columbia, SC, 29223
Occupant 050017, 2008, Twin Lakes Rd, Columbia, SC, 29209
Occupant 040018, 496, Kalmia Dr, Columbia, SC, 29205
Occupant 030019, 1168, Dothan Rd, Columbia, SC, 29210
Occupant 050020, 20, Foxtrail Dr, Columbia, SC, 29223
```

# Task: Address Matching

—TIGER/Line '95

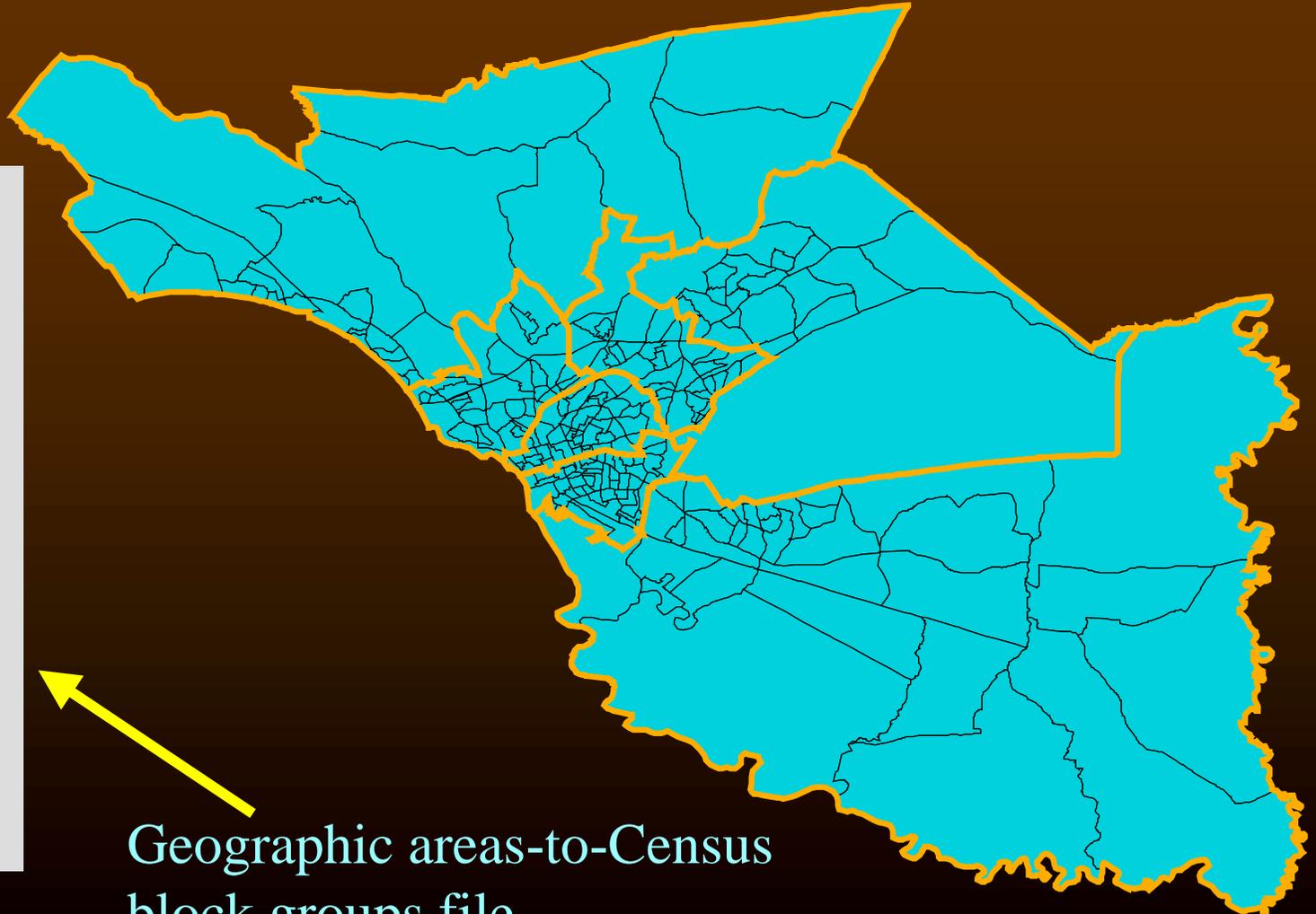
- Test address locations



# Task: Geographic Area Construction

TR90,BG,HD  
10100,1,1  
10100,3,7  
10302,1,2  
10200,1,3  
10100,2,2  
10200,3,1  
10200,2,5  
10302,2,4  
10303,1,6  
...  
...

Geographic areas-to-Census  
block groups file



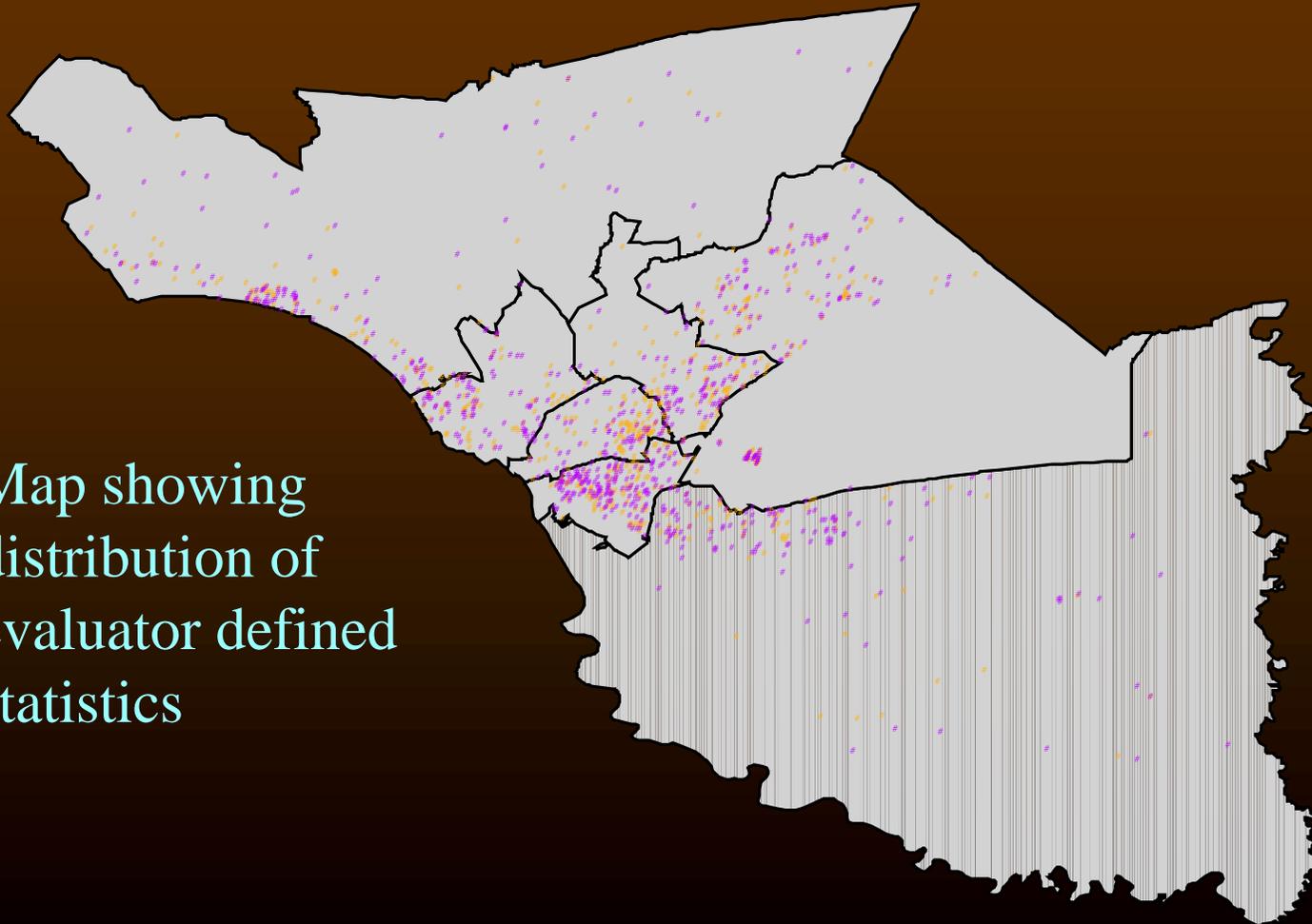
# Task: Rate Estimation

Portion of file used when testing for statistical significance of rates.

CaseID	CaseCode	StructureAddress	StreetID	Place	State	ZIP	Tract/BG	Longitude	Latitude
0001	11126187	1742	Overhill	Rd	Columbia	SC	29223, 0113053	-080.935138	+34.050065
0011	10224800	840	Delverton	Rd	Columbia	SC	29203, 0106003	-081.035133	+34.054533
0083	10132200	6902	Wakefield	Rd	Columbia	SC	29203, 0107024	-081.021340	+34.079266
0093	11223086	1033	Goodwin	Rd	Columbia	SC	29052, 0118003	-080.787628	+33.856137
0095	10234000	1847	High	St	Columbia	SC	29203, 0005002	-081.031325	+34.032483
0119	10122900	1882	Congress	Rd	Columbia	SC	29044, 0119021	-080.751279	+33.960006
0153	11226085	804	Vernon	St	Columbia	SC	29203, 0106002	-081.021482	+34.057650
0161	10122900	8124	Richard	St	Columbia	SC	29209, 0117022	-080.948396	+33.948944
0178	10326100	605	Goodwin	Way	Columbia	SC	29052, 0118003	-080.782940	+33.866493
0195	10324200	150	Glenshire	Dr	Columbia	SC	29203, 0107022	-081.001802	+34.084366
0204	10225900	102	Pointer	Dr	Columbia	SC	29061, 0119012	-080.876920	+33.973554
0245	11122584	633	Henry	St	Eastover	SC	29044, 0120003	-080.694695	+33.875474
0291	11326185	1038	Bluff	Rd	Columbia	SC	29201, 0117012	-081.020188	+33.971778
0304	10337300	271	Penrose	Dr	Columbia	SC	29203, 0107022	-081.001045	+34.078317
0320	11235388	9702	Garners Ferry	Rd	Columbia	SC	29061, 0119022	-080.842328	+33.951594
0365	11236682	1893	Long Shadow	Ln	Columbia	SC	29223, 0113054	-080.939446	+34.047632
0367	10222400	3895	Ardincaple	Dr	Columbia	SC	29203, 0004002	-081.057666	+34.035576

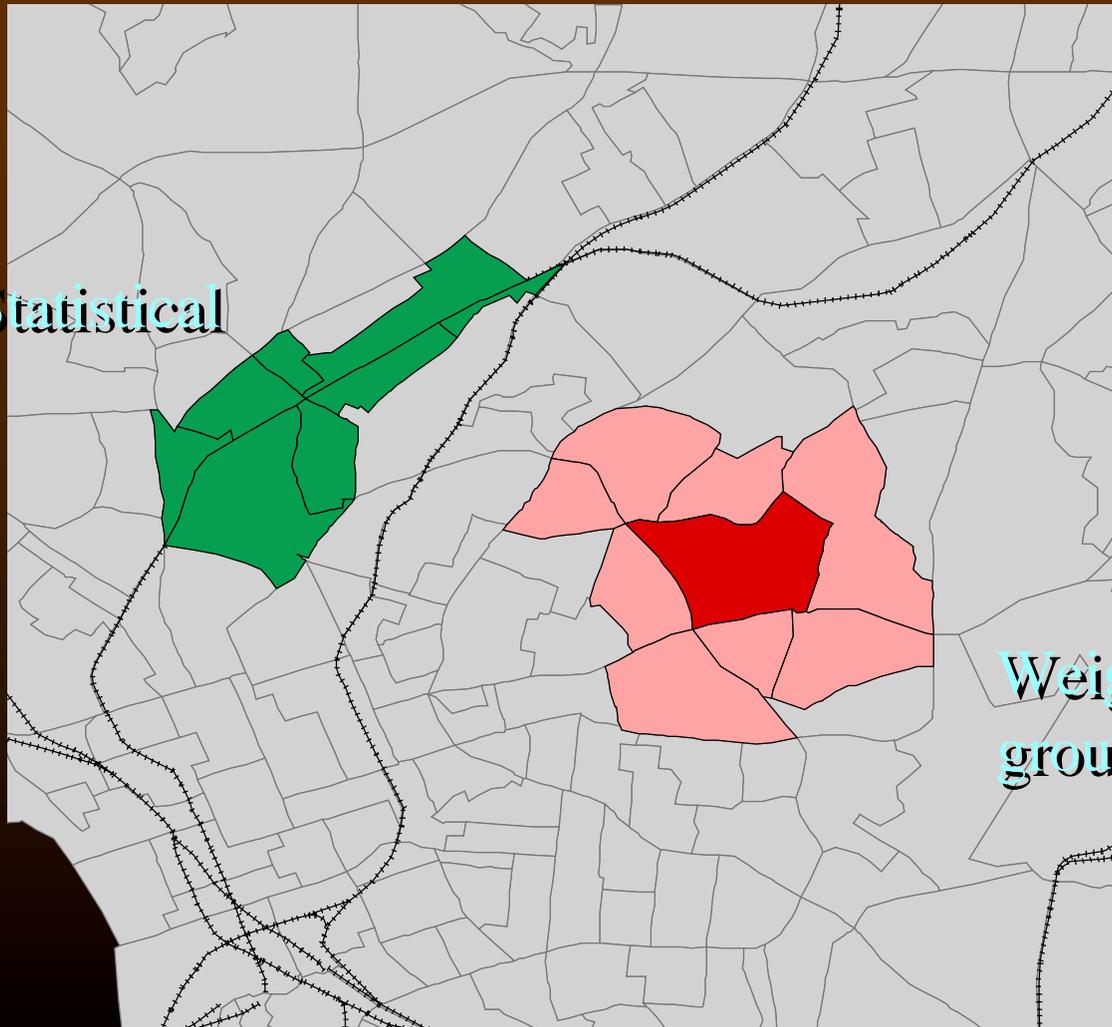
# Task: Rate Estimation

Map showing  
distribution of  
evaluator defined  
statistics



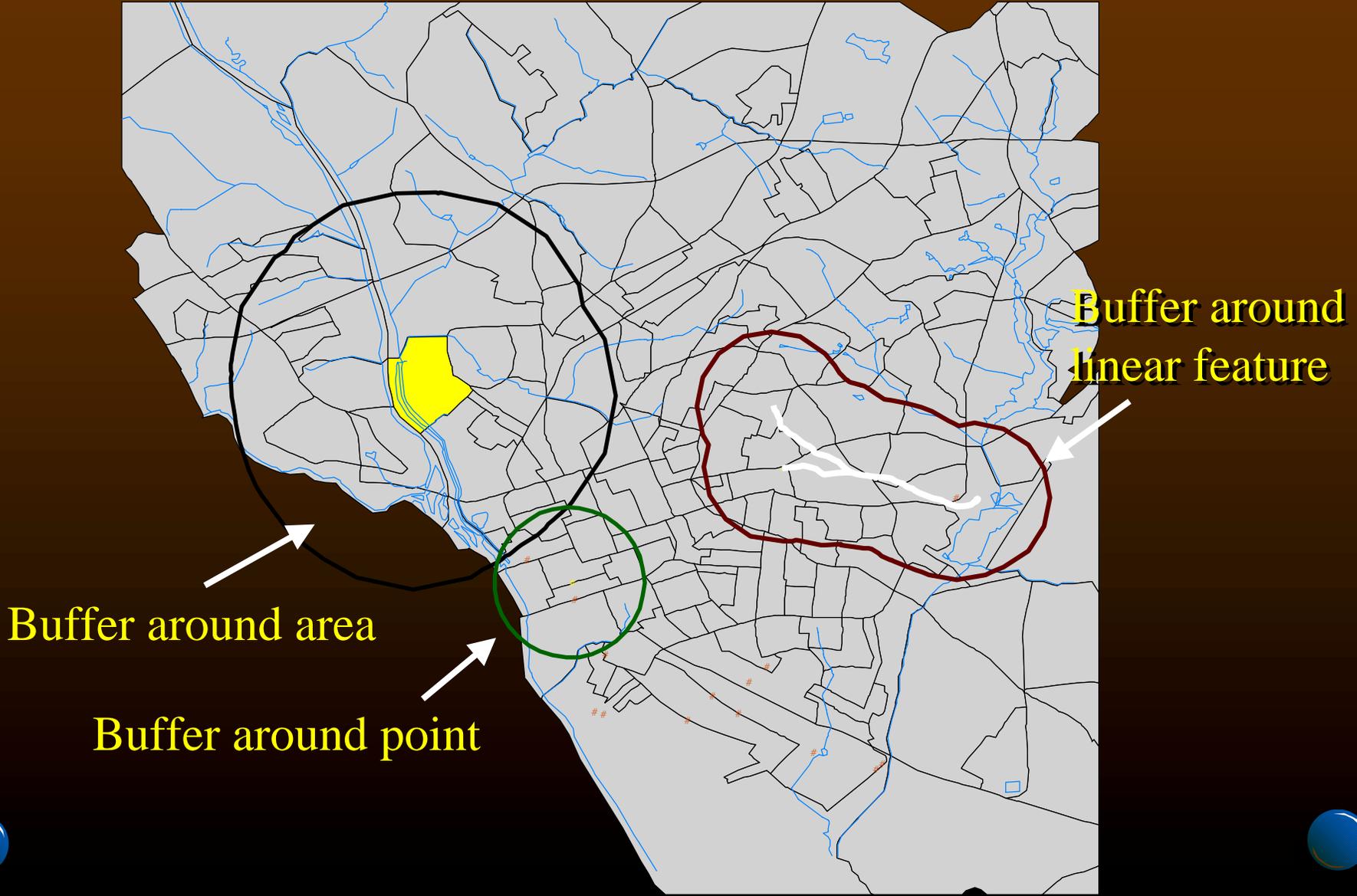
# Task: Spatial Cluster Detection

Testing for Statistical  
Significance

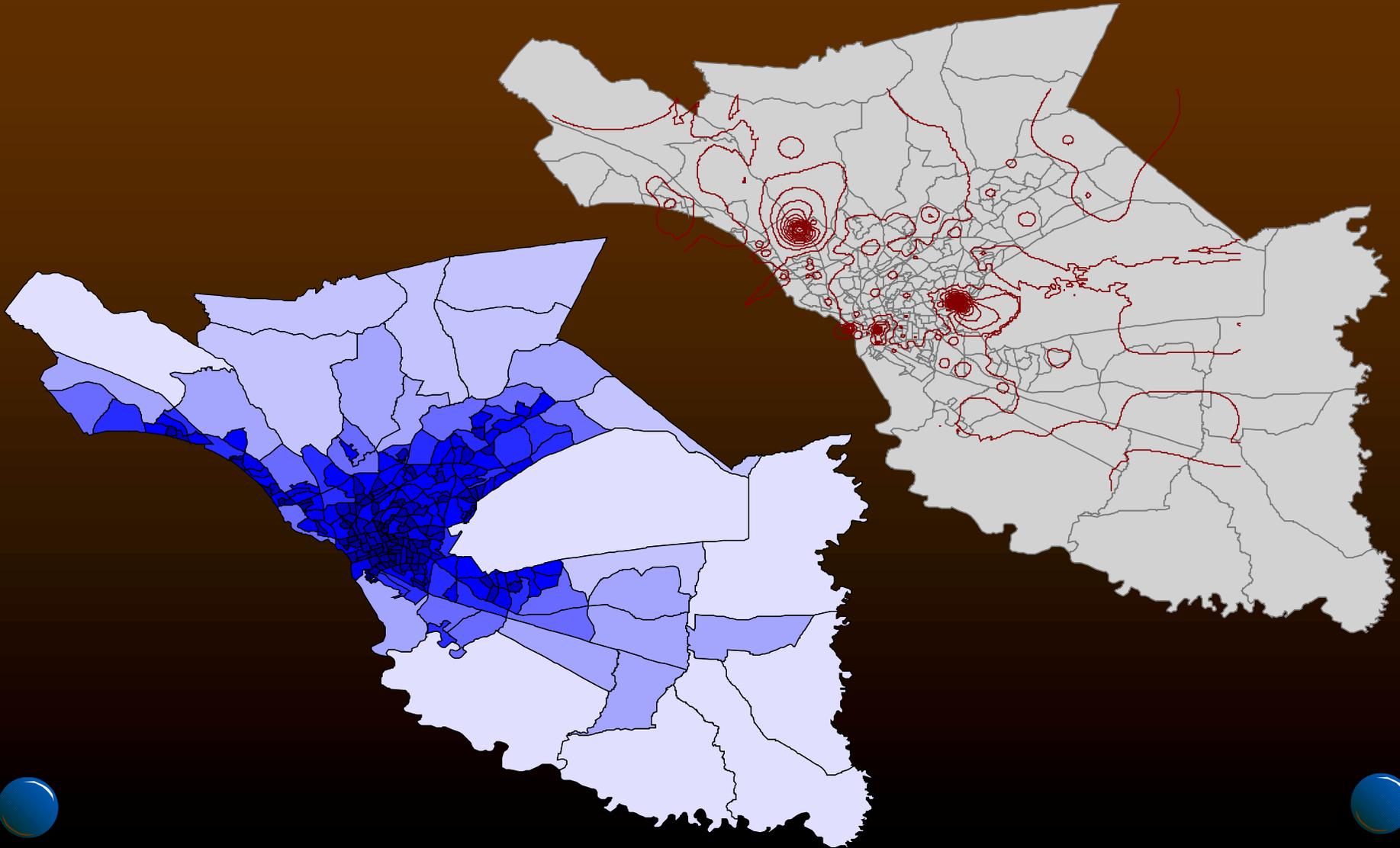


Weighted block  
groups

# Task: Buffering

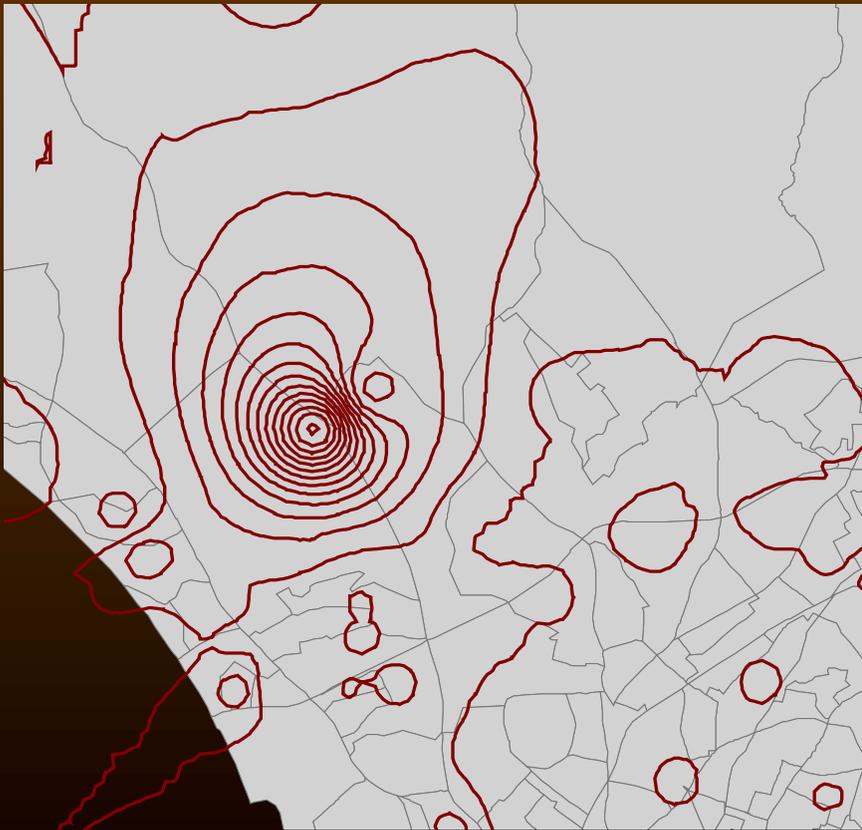


# Task: Contour/Choropleth Map Design

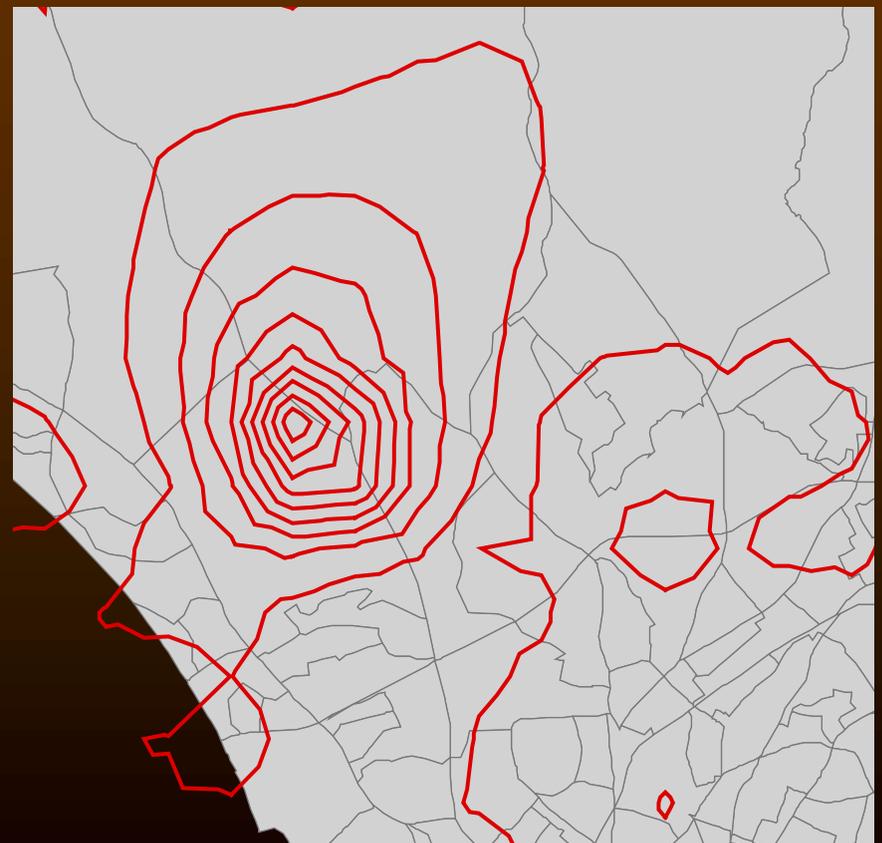


# Task: Spatial Smoothing

Initial surface



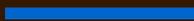
Smoothed surface



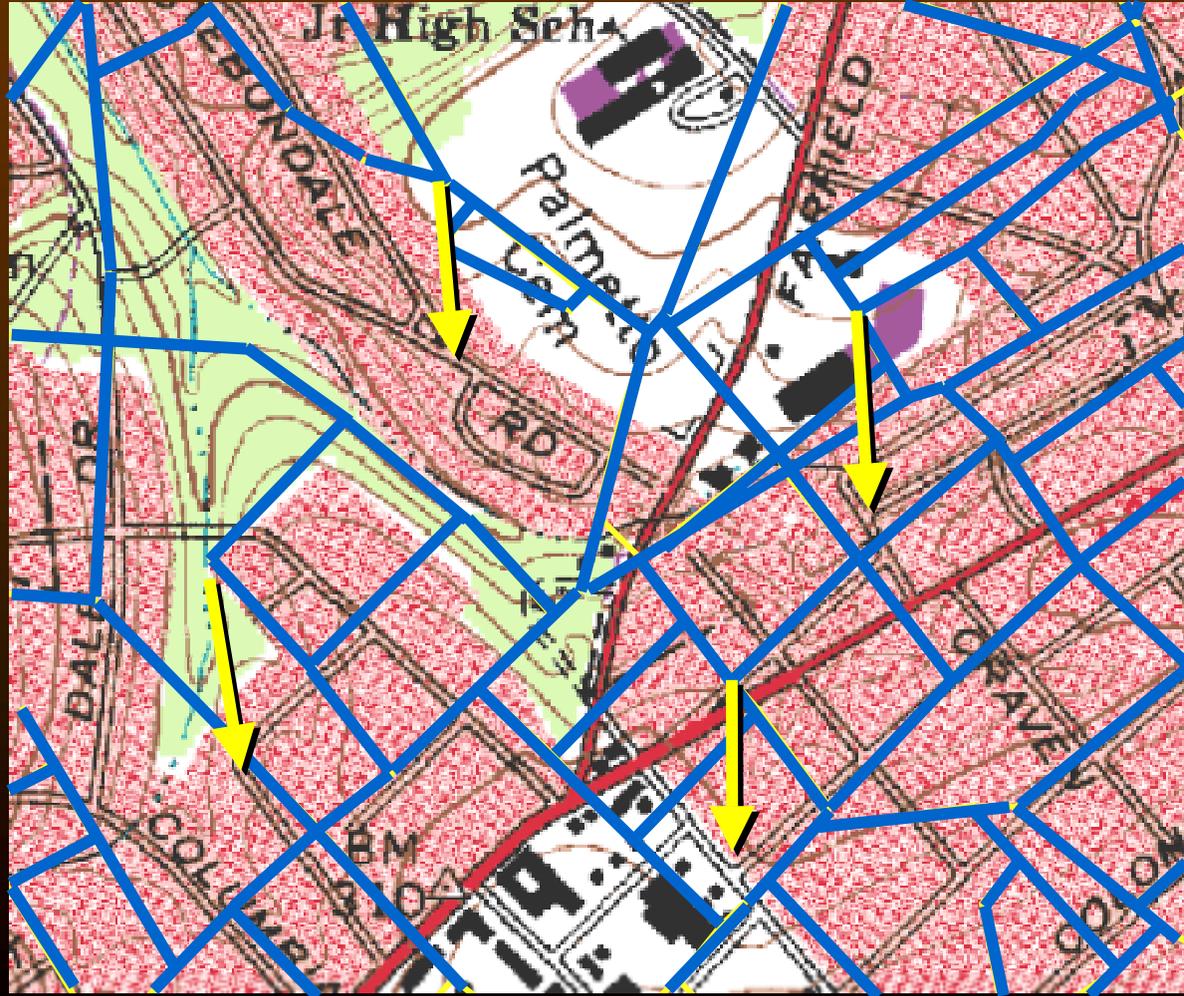
Contour interval remained constant

# Task: Map Overlay (dataset conflation)

TIGER/Line '95  
and a USGS  
7.5 min quad

TIGER data  
in blue 

Amount and  
direction of  
needed shift  

# Benchmark Test Development Issues

- ❑ Identification and selection of project teams
- ❑ Selecting test data types and geographic level
- ❑ Determining appropriate tasks for evaluation
- ❑ Example datasets for the tasks
- ❑ Example benchmark tasks
- ☑ Importance of a trial run
- ❑ Benchmark test administration
- ❑ Evaluating performance

# Trial Run

(Performed by test suite development team and evaluation team members)

## Purpose:

- ❑ Time the tasks
- ❑ Test task feasibility
- ❑ Make certain the test will evaluate the necessary abilities
- ❑ Develop test answer solutions
- ❑ Help evaluators understand proposer's approach and methodology

# Benchmark Test Development Issues

- ❑ Identification and selection of project teams
- ❑ Selecting test data types and geographic level
- ❑ Determining appropriate tasks for evaluation
- ❑ Example datasets for the tasks
- ❑ Example benchmark tasks
- ❑ Importance of a trial run
- ☑ Benchmark test administration
- ❑ Evaluating performance

# Test Administration

---

## Advance notification by letter

- Provides vendors a hint of items on upcoming test
- Helps for staffing and preparation for timed environment

## The Test Environment

- At proposer's site-designated by proposer
- Test may be over the course of multiple days
- Tasks are timed
- Oral explanations allowed, but hardcopy results should be required
- Members of evaluation team present, but not hanging over vendor's shoulder
- Opportunity to meet members of the team
- Opportunity for vendor to highlight work related to project

# Benchmark Test Development Issues

- Identification and selection of project teams
- Selecting test data types and geographic level
- Determining appropriate tasks for evaluation
- Example datasets for the tasks
- Example benchmark tasks
- Importance of a trial run
- Benchmark test administration
- Evaluating performance

# Evaluating Proposer Performance

---

- ❑ Where      Test site and offsite
- ❑ Who        Voting members and technical support
- ❑ How        Evaluations done individually and then discussed as a group
- ❑ When       Test site and later in a meeting to make recommendation
- ❑ What to evaluate    Hard copy results only

## Contact information:

Dr. Ellen Heineman  
National Cancer Institute

[heinemae@epndce.nci.nih.gov](mailto:heinemae@epndce.nci.nih.gov)

Frederick R. Broome  
Geospatial Research and Standards Staff  
Geography Division  
US Bureau of the Census

301-457-1056  
[fbroome@census.gov](mailto:fbroome@census.gov)