

The Spatial Context of Health Disparities:



University of Connecticut

Findings from the UConn-DPH Geocoding Collaborative

Wednesday, December 10, 2008

1:00 to 4:00 PM

The Lyceum

Hartford, Connecticut





A Protocol for Geocoding CT Mortality Databases

By Peter Hayward, Brandon Cramer, & Lloyd Mueller

Acknowledgements

UCONN—DPH GEOCODING COLLABORATIVE



UCONN—Professors Jeffery Osleeb and Alexander Vias

DPH—Margaret Hynes and Karyn Backus



Presentation Outline

- CT Mortality Data
- What is Geocoding?
- The UCONN—DPH Collaborative
- Process of Geocoding CT Mortality Data
- Results and Conclusions

CT Mortality Data



CT Mortality Data

- Connecticut mortality records for each individual death (stripped of identifiers) across a specified number of years.

- Variables Can Include
 - Last known street address, town, and zip-code
 - Year of death
 - Age
 - Sex
 - Race
 - Cause of death
 - etc., etc., etc.

CT Mortality Data

Attributes of deaths8504fwDemogr_foruconn

OID	DTHYR	SFII	SEX	HISPANIC	AGE	DRACE	XCAUSEP	TWIRES	ZIP	Origstreet
46342	1989	198918673	1	0	60	1	0	15	06605	56 HAZELWOOD AV
46343	1988	198814888	2	0	72	1	0	15	06605	35 HEMLOCK ST
46344	1986	198606779	2	0	79	1	0	15	06605	85 HEMLOCK ST
46345	1988	198827912	1	0	82	1	0	15	06605	23 HOMESTEAD AV
46346	1987	198713816	1	0	62	1	0	15	06605	64 HOMESTEAD AV
46347	1986	198605046	1	0	63	1	0	15	06605	7 HOMESTEAD AV
46348	1988	198802885	2	0	81	1	0	15	06605	121 HOPE ST
46349	1986	198612415	2	0	91	1	0	15	06605	45 HOPE ST
46350	1985	198522421	2	0	69	1	0	15	06605	80 HOPE ST
46351	1990	199003008	1	0	19	2	0	15	06605	96 HOPE ST
46352	1990	199005360	1	1	43	1	0	15	06605	1018 HOWARD AV
46353	1985	198523511	2	0	37	2	0	15	06605	1021 HOWARD AV
46354	1990	199012093	1	0	73	1	0	15	06605	1023 HOWARD AV
46355	1986	198612382	1	0	56	2	0	15	06605	1027 HOWARD AV
46356	1991	199100259	2	0	78	1	0	15	06605	1047 HOWARD AV
46357	1986	198618492	1	0	68	1	0	15	06605	1159 HOWARD AV
46358	1988	198800308	1	5	53	1	0	15	06605	1274 HOWARD AV
46359	1985	198502680	1	0	71	1	0	15	06605	1281 HOWARD AV
46360	1989	198902611	2	0	90	2	0	15	06605	1281 HOWARD AV



CT Mortality Data

- Problem
 - Can identify who, when, and how
 - Need to rectify the question of where

- Why Do We Need to Know Where
 - A spatial perspective provides insight into the health attributes of certain segments of society in distinct locations

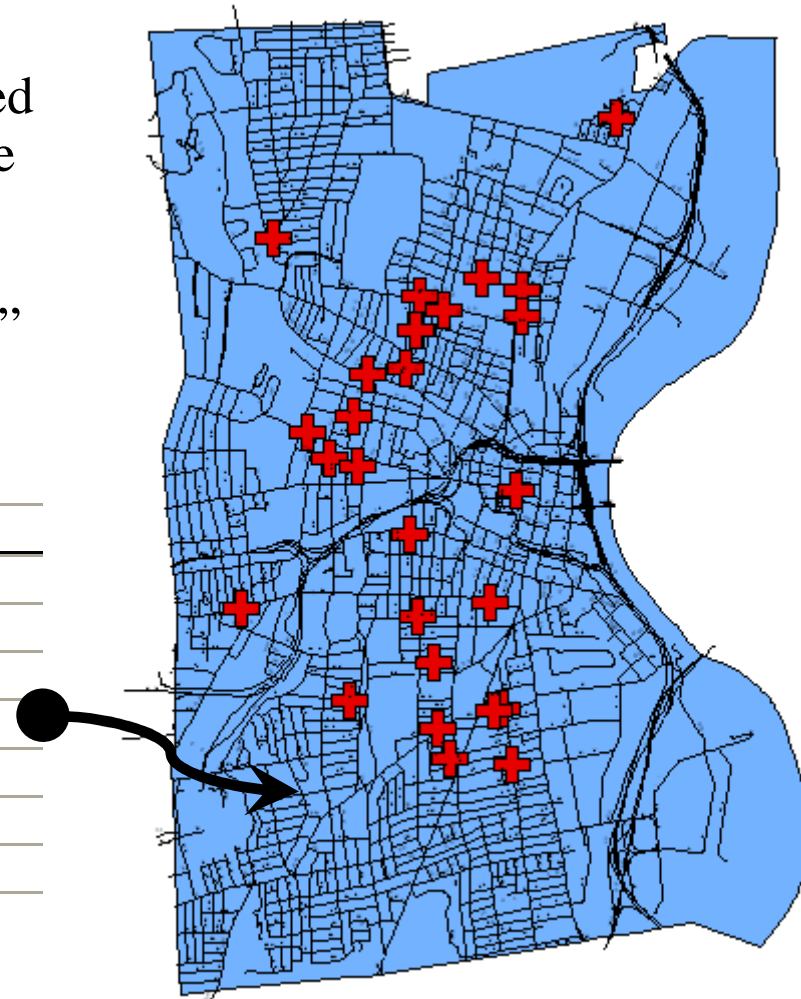
- Solution
 - Geographic Information Systems (GIS) and the process of geocoding

What is Geocoding?

What is Geocoding?

- Formal Definition
 - “The matching of a location stored in a table to a spatial point feature based on a reference spatial data layer; most often applies to converting addresses to locations”
(Price 2006)

Acc_Number	Address
1000	227 Lawrence Street, Hartford
1001	370 Capitol Avenue, Hartford
1002	45 Park Avenue, Hartford
1003	5113 Main Street, Hartford
1004	32 Main Street Hartford
1005	570 Whitney Street, Hartford
1006	77 Jefferson Avenue, Hartford
1007	2120 Vine Street. Hartford





The UCONN—DPH Collaborative



Purpose

- ❑ To examine the process of geocoding 1985 - 2004 CT mortality data using ArcGIS 9.x for the purpose of creating mappable points
- ❑ To outline the problems typically encountered in the geocoding process and to treat the problems using innovative solutions
- ❑ To measure the “geocoding success rate” by a number of variables



Process of Geocoding
1985 - 2004 CT Mortality Data

Necessary Component: CT Mortality Data

Attributes of deaths8504fwDemogr_foruconn

OID	DTHYR	SFNI	SEX	HISPANIC	AGE	DRACE	XCAUSEP	TWIRES	ZIP	Origstreet
46342	1989	198918673	1	0	60	1	0	15	06605	56 HAZELWOOD AV
46343	1988	198814888	2	0	72	1	0	15	06605	35 HEMLOCK ST
46344	1986	198606779	2	0	79	1	0	15	06605	85 HEMLOCK ST
46345	1988	198827912	1	0	82	1	0	15	06605	23 HOMESTEAD AV
46346	1987	198713816	1	0	62	1	0	15	06605	64 HOMESTEAD AV
46347	1986	198605046	1	0	63	1	0	15	06605	7 HOMESTEAD AV
46348	1988	198802885	2	0	81	1	0	15	06605	121 HOPE ST
46349	1986	198612415	2	0	91	1	0	15	06605	45 HOPE ST
46350	1985	198522421	2	0	69	1	0	15	06605	80 HOPE ST
46351	1990	199003008	1	0	19	2	0	15	06605	96 HOPE ST
46352	1990	199005360	1	1	43	1	0	15	06605	1018 HOWARD AV
46353	1985	198523511	2	0	37	2	0	15	06605	1021 HOWARD AV
46354	1990	199012093	1	0	73	1	0	15	06605	1023 HOWARD AV
46355	1986	198612382	1	0	56	2	0	15	06605	1027 HOWARD AV
46356	1991	199100259	2	0	78	1	0	15	06605	1047 HOWARD AV
46357	1986	198618492	1	0	68	1	0	15	06605	1159 HOWARD AV
46358	1988	198800308	1	5	53	1	0	15	06605	1274 HOWARD AV
46359	1985	198502680	1	0	71	1	0	15	06605	1281 HOWARD AV
46360	1989	198902611	2	0	90	2	0	15	06605	1281 HOWARD AV

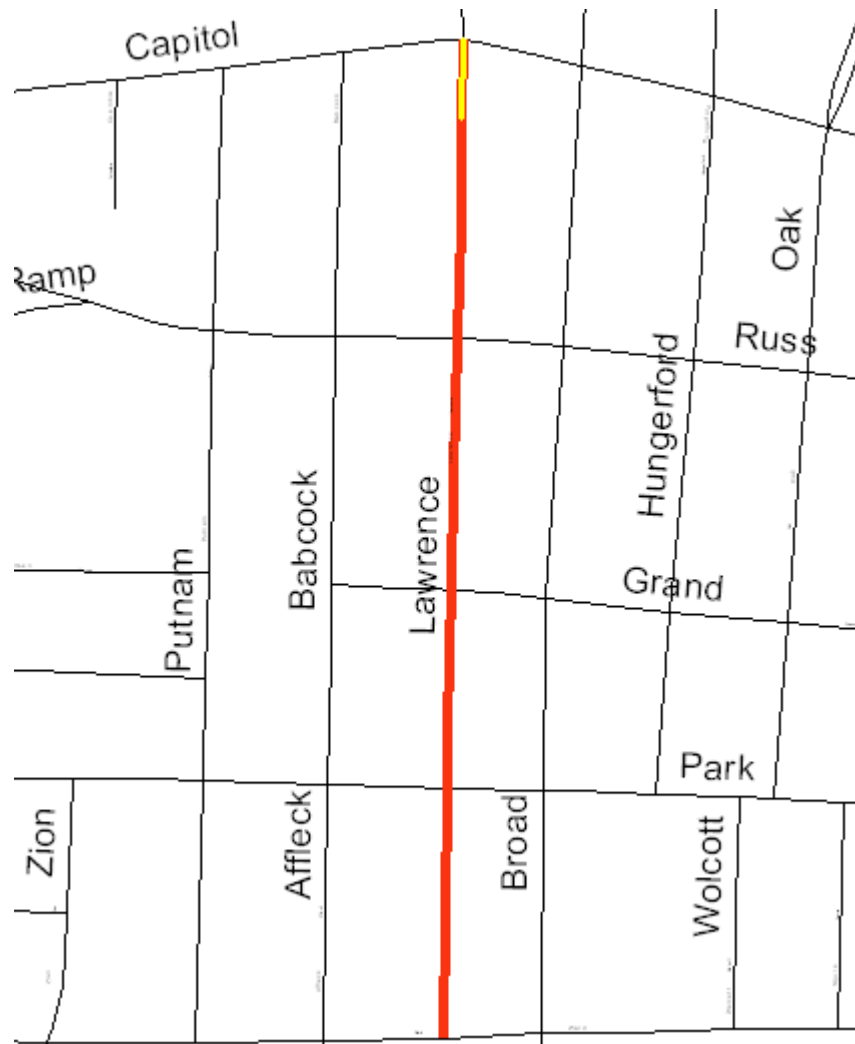
Necessary Component: Reference Database

	Shape *	DYHA
18	Polyline	4!
17	Polyline	4!
16	Polyline	4!
15	Polyline	4!
16	Polyline	4!
15	Polyline	4!
14	Polyline	4!
28	Polyline	4!
12	Polyline	4!
30	Polyline	4!
10	Polyline	4!
39	Polyline	4!
38	Polyline	4!
37	Polyline	4!
36	Polyline	4!
35	Polyline	4!
34	Polyline	4!
33	Polyline	4!
32	Polyline	4!
2	Polyline	4!
13	Polyline	4!
3	Polyline	4!
10	Polyline	4!
39	Polyline	4!
38	Polyline	4!
37	Polyline	4!



TYPE	SU
St	
St	
<Null>	
Ave	
St	
St	
St	
St	
St	
Ter	
St	
Ter	
St	
Ter	
St	
St	
St	
St	
Ter	
Hts	
St	
Ave	
St	
Ave	
St	

Necessary Component: Reference Database



	L_F_ADD	L_T_ADD	R_F_ADD	R_T_ADD	NAME	TYPE
§ 1		99	2	56	Lawrence	St
§ 101		105	58	74	Lawrence	St
§ 107		115	76	102	Lawrence	St
r 117		121	104	120	Lawrence	St
0 123		199	122	198	Lawrence	St
r 201		273	200	272	Lawrence	St
0 275		299	274	298	Lawrence	St

Necessary Component: Address Locator

- A ArcGIS tool that allows user to input the reference database and set parameters for searching for addresses

The screenshot shows the 'Address Locator Properties' dialog box with several sections and callout boxes:

- Naming Options:** A callout box asks, "Is the address field called 'address', 'street', or something else?". The dialog shows 'Street' and 'Zone' in the 'Input Address Fields' list, and 'Address' and 'Street' in the 'is recognized if it is named:' list.
- Matching Options:** A callout box states, "The lower the values, the more likely it is to receive a successful match." The dialog shows sliders for 'Spelling sensitivity' (60), 'Minimum candidate score' (10), and 'Minimum match score' (80).
- Output Options:** A callout box asks, "What units the output addresses data points will refer to (feet, meters, yards), whether or not to include latitude/longitude?". The dialog shows 'Side offset' (20) and 'End offset' (3%) with a unit dropdown set to 'Reference data units'. It also has checkboxes for 'Match if candidates tie', 'X and Y coordinates', 'Reference data ID', 'Standardized address', and 'Percent along'.



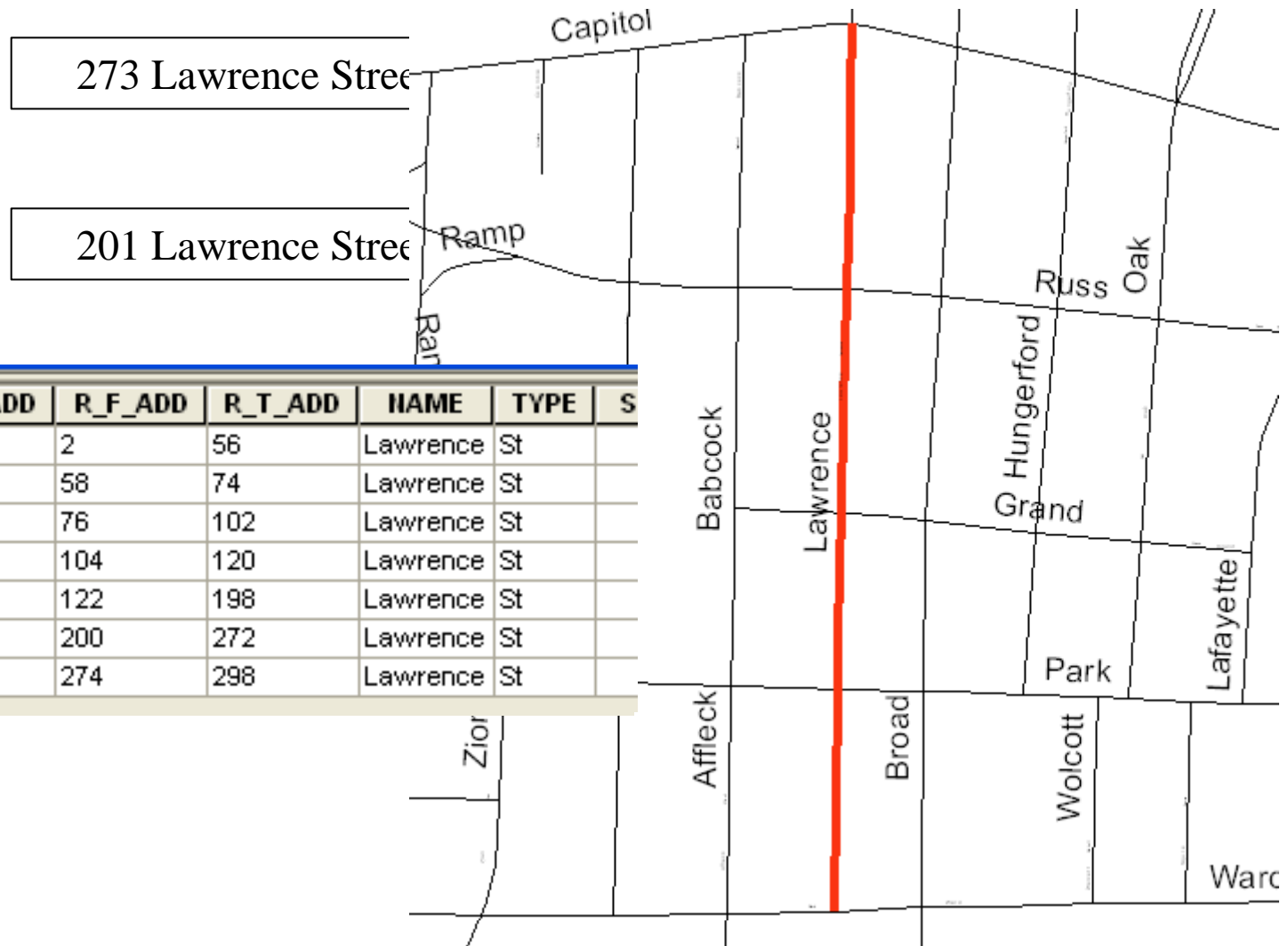
Geocoding: Linear Referencing

- Given the necessary components, geocoding took place through the process of linear referencing
- Linear Referencing is the process of “using distance measures to locate events along a line”

(ESRI 2008)

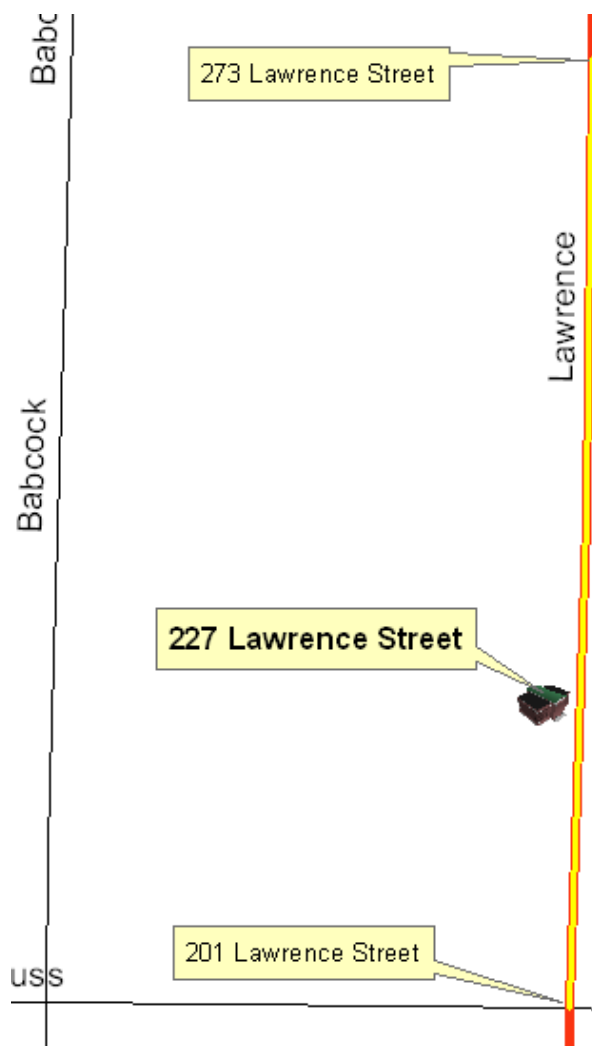
- Example: Locate 227 Lawrence Street

Locate 227 Lawrence Street



ID	L_F_ADD	L_T_ADD	R_F_ADD	R_T_ADD	NAME	TYPE	S
6	1	99	2	56	Lawrence St		
3	101	105	58	74	Lawrence St		
6	107	115	76	102	Lawrence St		
7	117	121	104	120	Lawrence St		
0	123	199	122	198	Lawrence St		
7	201	273	200	272	Lawrence St		
0	275	299	274	298	Lawrence St		

Locate 227 Lawrence Street





Geocoding Success Rate

- The collaborative attempted to geocode 578,860 records (1985 – 2004 mortality data)
- Initial results indicated a geocoding success rate of 90 percent

$$\text{Geocoding Success Rate} = \frac{\text{Matched Records}}{\text{Total Records}} \times 100$$



Identifying Problems

- To improve upon the geocoding success rate, the collaborative summarized the major problems by analyzing the unmatched CT mortality address data

- Problem addresses were consolidated using the “Summarize” function in ArcGIS
 - Corollary to “Frequency” in SPSS, SAS



Problems

- ❑ Missing address information in CT Mortality Data
- ❑ Spelling mistakes in CT Mortality Data
- ❑ Incorrect street designations in CT Mortality Data

Problem: Missing Address Information

XCAUSEP	TWIRES	ZIP	Origstreet
4	0		Covenant Village
0	0		Covenant Village
0	0		Covenant Village
0	0		Covenant Village
0	0		Covenant Village
0	0		Covenant Village



Problem: Spelling Mistakes

XCAUSEP	TWIREs	ZIP	Origstreet
0	0		75 Wills Street
0	0		75 Wills Street
0	0		75 Wills Street
0	0		220 Seamor Street
0	0		220 Seamor Street
0	0		220 Seamor Street
0	0		39 Laffayet Street
0	0		39 Laffavet Street



Problem: Incorrect Street Designations

XCAUSEP	TWIRRES	ZIP	Origstreet
0	0		52 West St.
0	0		52 West St.
0	0		52 West St.
0	0		52 West St.
0	0		52 West St.
0	0		52 West St.
0	0		52 West St.



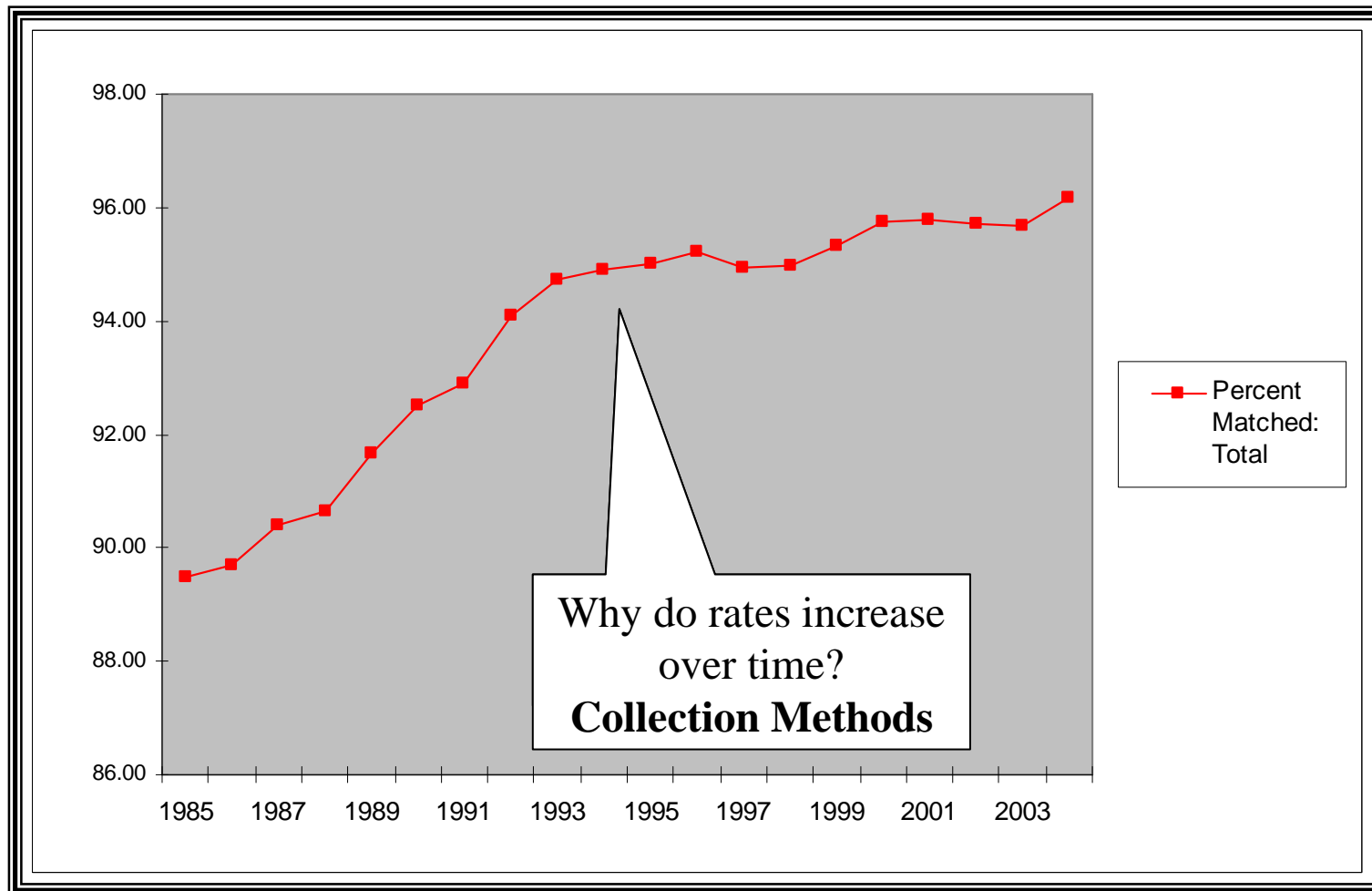
Results and Conclusions



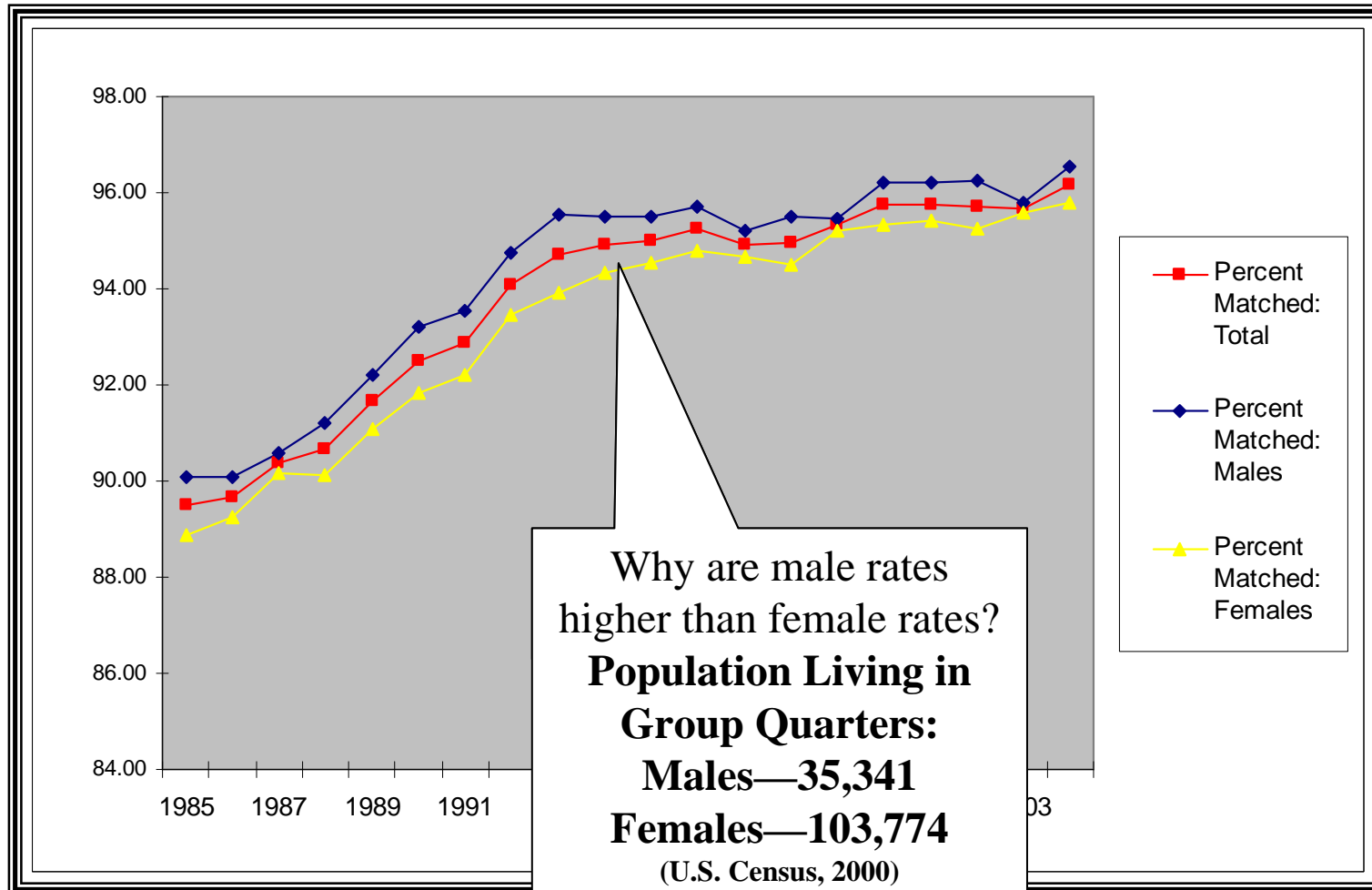
Geocoding Success Rate

- ❑ Of the 578,860 CT Mortality Data records, 543,111 were successfully matched while 35,749 were left unmatched
- ❑ Improved the geocoding success rate from 90 percent to nearly 94 percent
- ❑ Those with missing information accounted for 54 percent of all unmatched records

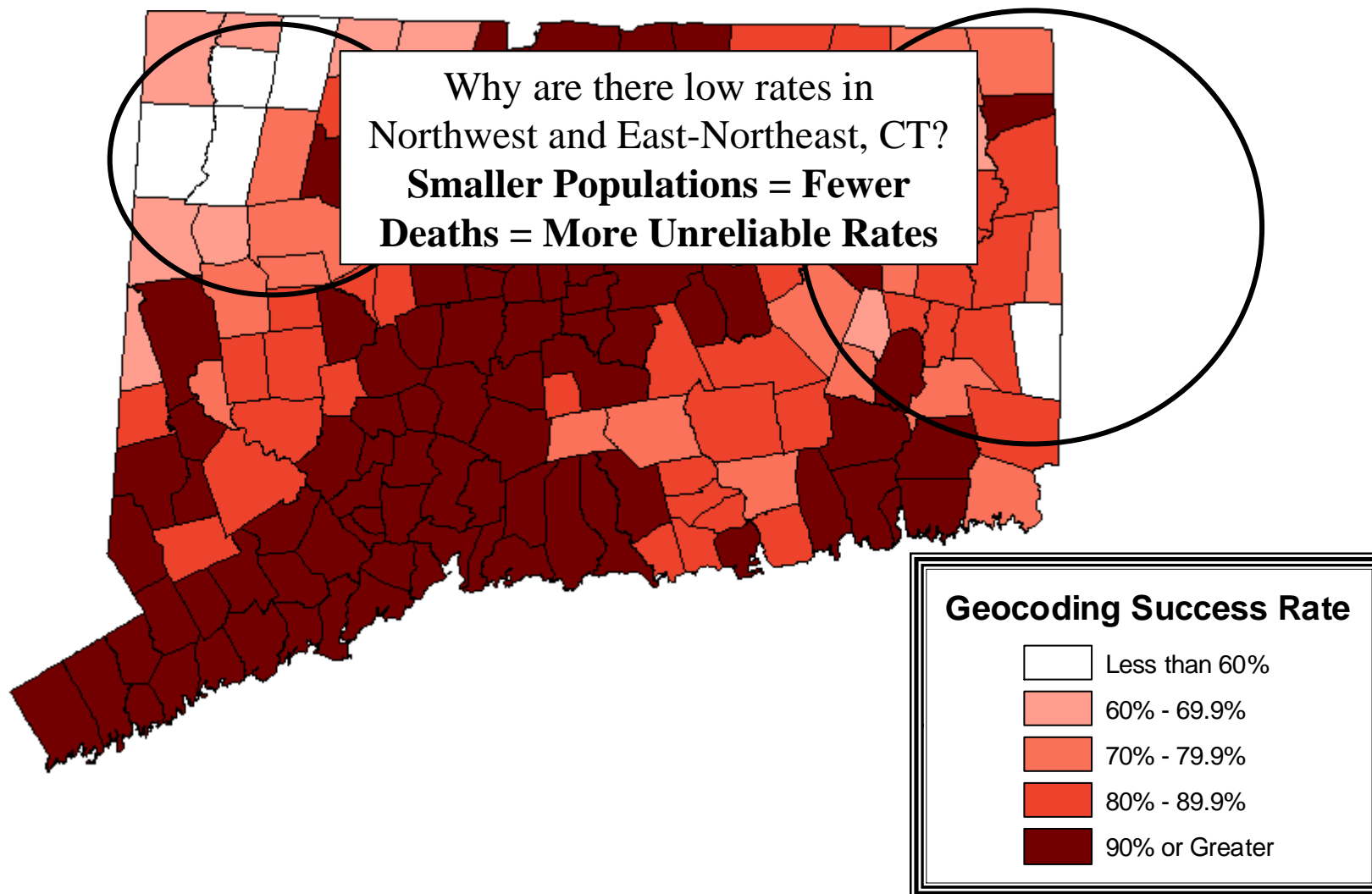
Geocoding Success Rate by Year



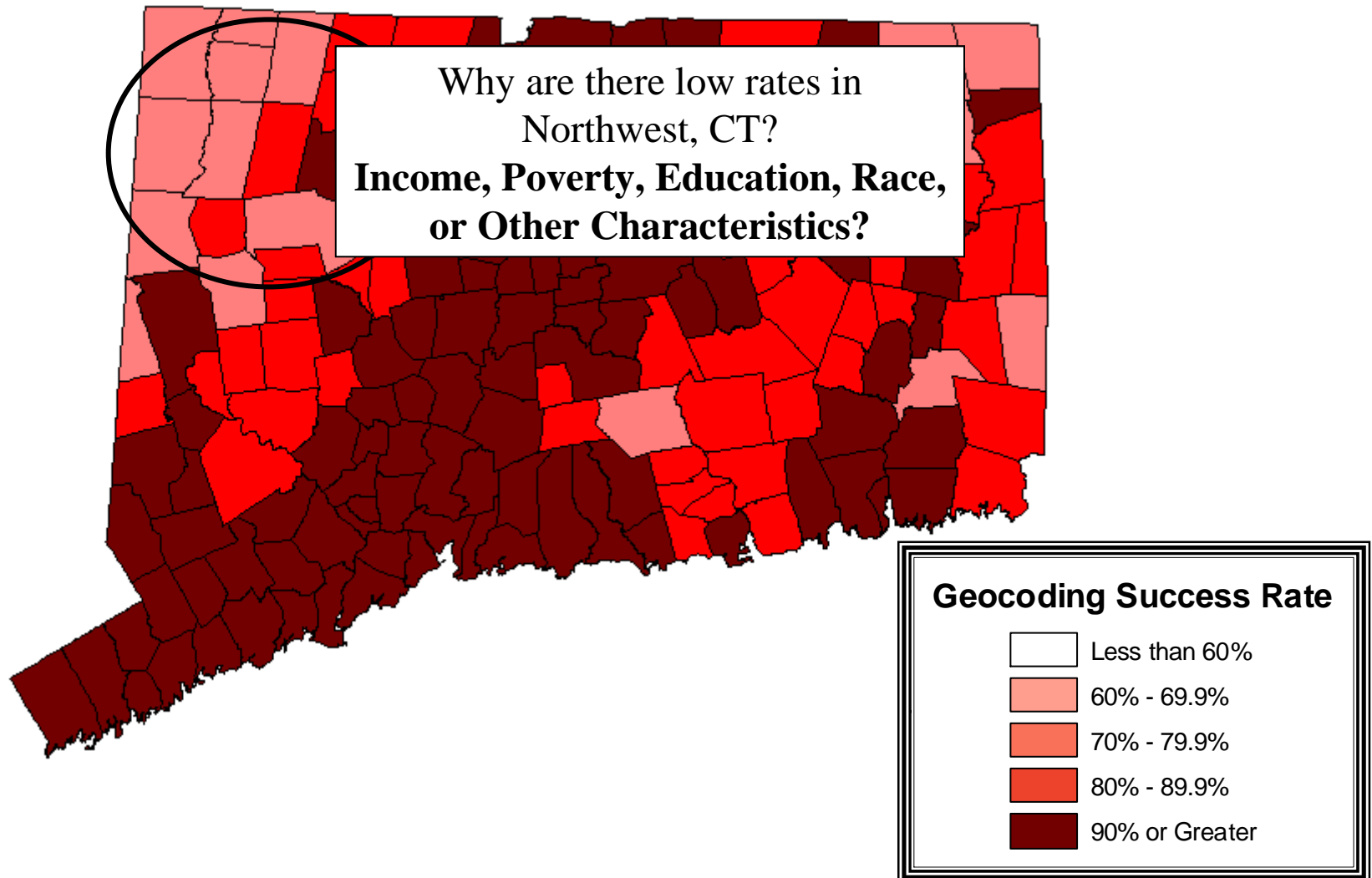
Geocoding Success Rate by Year & Sex



Geocoding Success Rate by Town



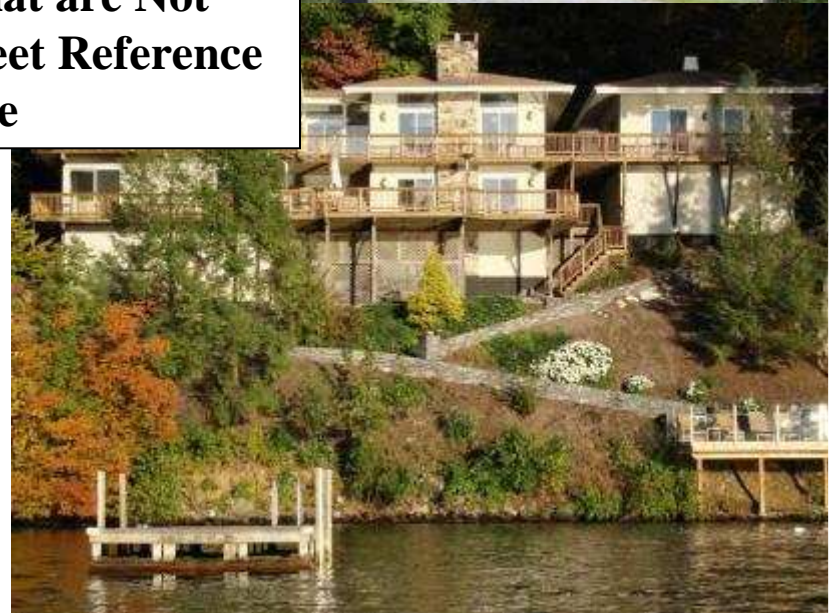
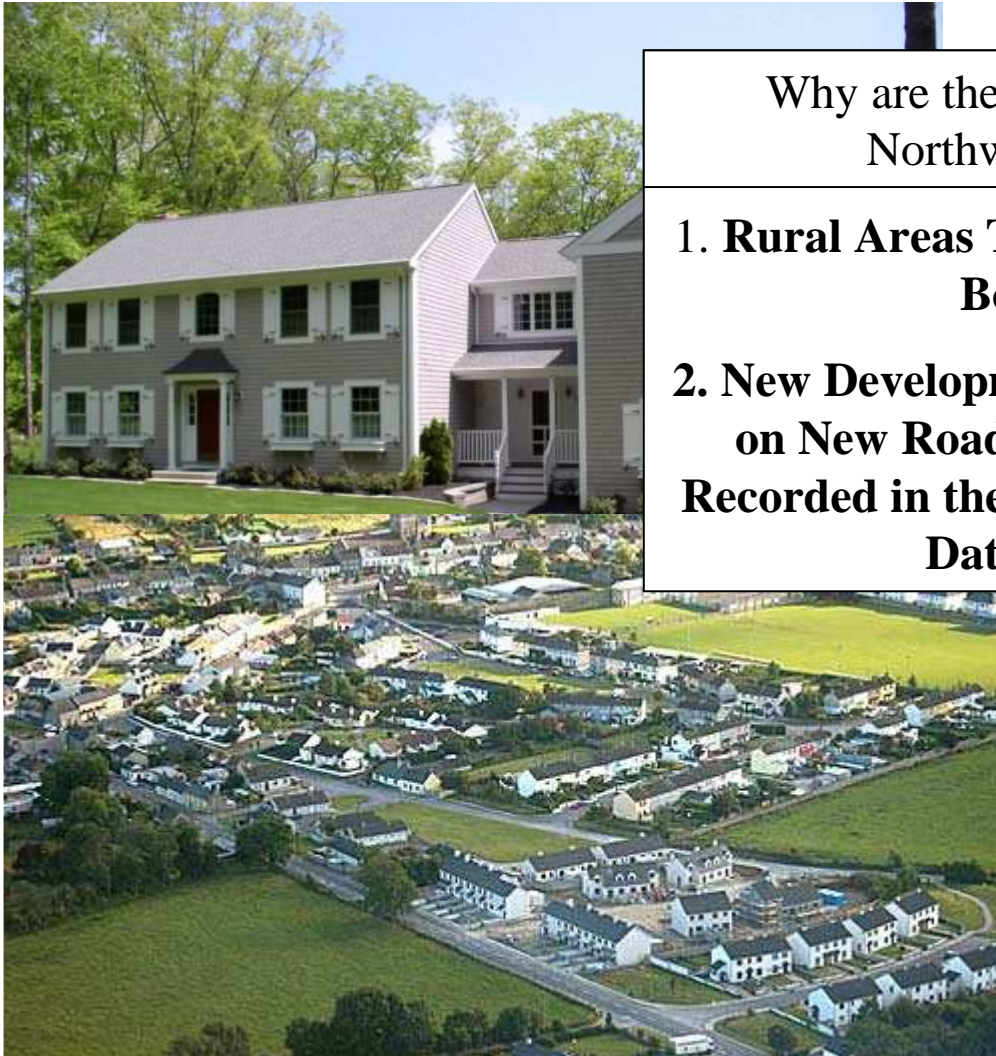
Geocoding Success Rate by Town (Adj.)



Geocoding Problems: Northwest, CT

Why are there low rates in Northwest, CT?

- 1. Rural Areas Tend to Have P.O. Boxes**
- 2. New Developments are Located on New Roads That are Not Recorded in the Street Reference Database**





Conclusions

- 1985 - 2004 CT mortality data was geocoded using ArcGIS 9.x for the purpose of creating mappable points
- Some problems encountered in the geocoding process were outlined and treated using innovative solutions. This increased the number of mappable points.
- Geocoding success rate can and should be measured by a number of variables and related to other factors.

Acknowledgements

UCONN—DPH GEOCODING COLLABORATIVE



UCONN—Professors Jeffery Osleeb and Alexander Vias

DPH—Margaret Hynes and Karyn Backus