

Improving the Quality of Geocoded Data

NCCCP & NPCR Conference

April 15, 2009

Kevin C. Ward, PhD, CTR

Georgia Center for Cancer Statistics



EMORY
UNIVERSITY

Census Geography

Geographic Unit

State

County

Census Tract (average 4,000 persons)

Block Group (average 1,000 persons)

Latitude/Longitude (point data)

ZIP Code (average 30,000 persons)

(Can cross state, county, tract and block group boundaries)

Geocoding Definition

The process of creating a spatial representation for a location (census tract, lat/long coordinates) from a textual description of the location (address).

3020 Frazier Wood Road Dec, GA 30033 EZ-Locate returned a segment match .		
Matched Address	Location (WGS-84)	2000 Census
3020 FRAZIER WOODS RD DECATUR, GA 30033	<u>Decimal Degrees</u> <u>Deg:Min:Sec</u> Lat: 33.830656 33:49:50.362N Lon: -84.265918 84:15:57.305W	FIPS State...: 13 FIPS County...: 089 Tract.....:0217.04 Block.....: 2011 FIPS Place...: FIPS MCD.....: 93096 MSA.....: 0520 CBSA.....: 12060 Metro.....: 0 Micro.....: 0
Standard Address	Postal	
3020 FRAZIER WOODS RD DECATUR,GA 30033-1843	Carrier.....: C014	

Uses of Geocoded Data

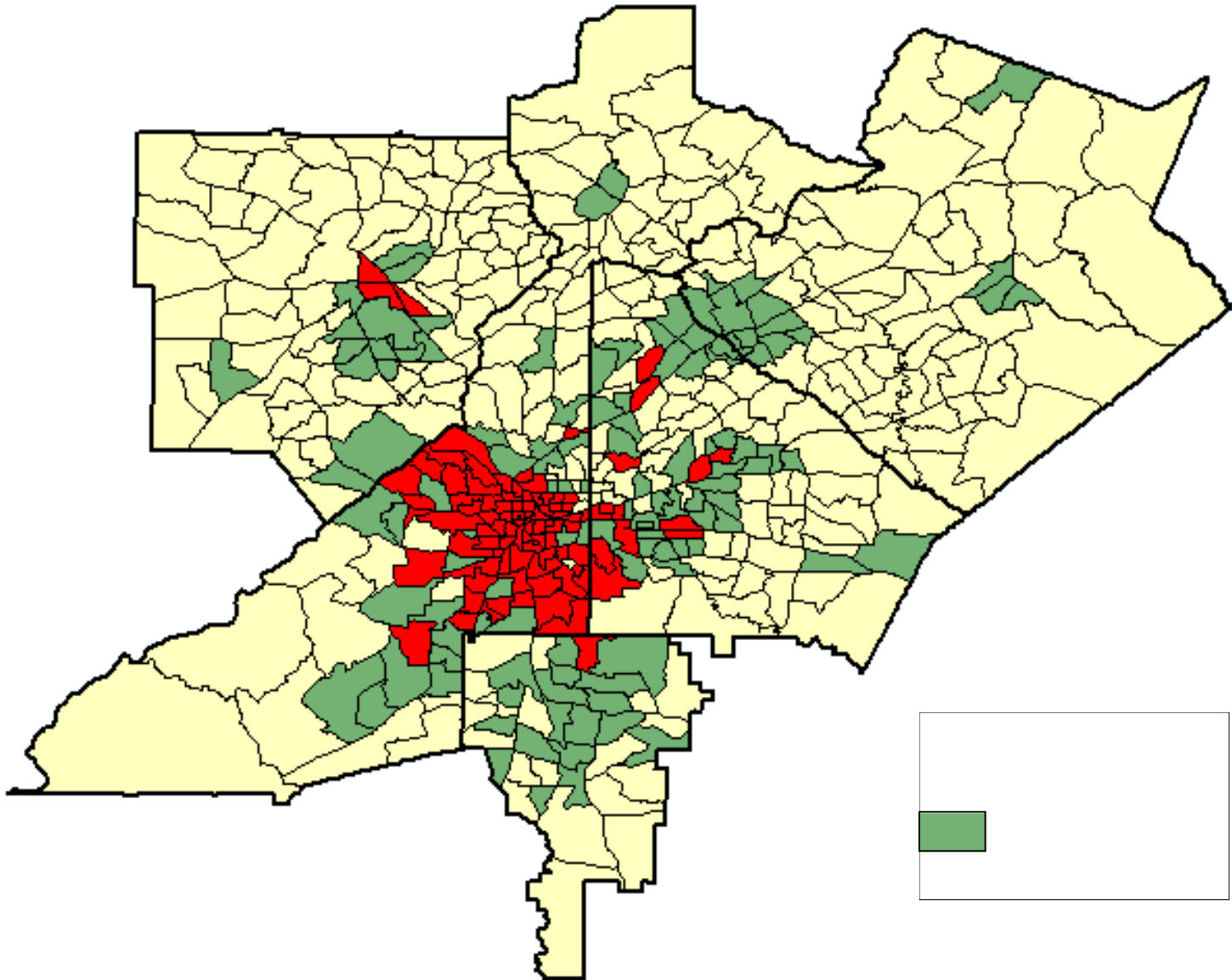
Area-based measures of socioeconomic status

3020 Frazier Wood Road Dec, GA 30033 EZ-Locate returned a segment match .		
Matched Address	Location (WGS-84)	2000 Census
3020 FRAZIER WOODS RD DECATUR, GA 30033	<u>Decimal Degrees</u> <u>Deg:Min:Sec</u> Lat: 33.830656 33:49:50.362N Lon: -84.265918 84:15:57.305W	FIPS State...: 13 FIPS County...: 089 Tract.....:0217.04 Block.....: 2011 FIPS Place...: FIPS MCD.....: 93096 MSA.....: 0520 CBSA.....: 12060 Metro.....: 0 Micro.....: 0
Standard Address	Postal	
3020 FRAZIER WOODS RD DECATUR,GA 30033-1843	Carrier.....: C014	

Source: U.S. Census Bureau, Census 2000 Summary File 3

Geography Identifier	Geography Identifier	Geog Summary Level	Geography	Individuals for whom poverty status is determined	Number; Below poverty level	Percent below poverty level
14000US13 089021704	13089021704	140	Census Tract 0217.04	5113	135	2.6

Maps by Poverty



Georgia Mortality by Poverty (1999-2001)

White Males				
Poverty Level				
Age	Low (0-9.9%)	Middle (10-19.9%)	High (20+%)	National
40-44 years	210.7	445.0	512.1	277.7
45-49 years	316.9	596.0	860.1	411.5
50-54 years	415.1	755.1	1456.4	583.2
55-59 years	681.0	1177.8	1827.0	922.1
60-64 years	1218.2	1909.3	2917.3	1457.1
65-69 years	2009.3	2811.1	3744.0	2299.3
70-74 years	3354.9	4223.5	6138.7	3600.5
75-79 years	5218.0	6468.8	7617.1	5619.6
80-84 years	9193.0	10295.6	12600.2	8987.6

Uses of Geocoded Data

Explore associations of distance from cancer patient's residence to diagnosis and/or treatment facilities

NAACCR research project with Komen Foundation

Utilize “Shortest Path Algorithm” to measure driving times and distances.

Analyze whether longer driving time between breast cancer patient's residence and diagnosis facility contributes to later stage.

Calculate Driving Distances / Times (5.4 miles, 12 minutes)

A 1364 Clifton Rd NE, Atlanta, GA 30322

B 3020 Frazier Woods Road Decatur Ga 30033

[Add Destination - Show options](#)

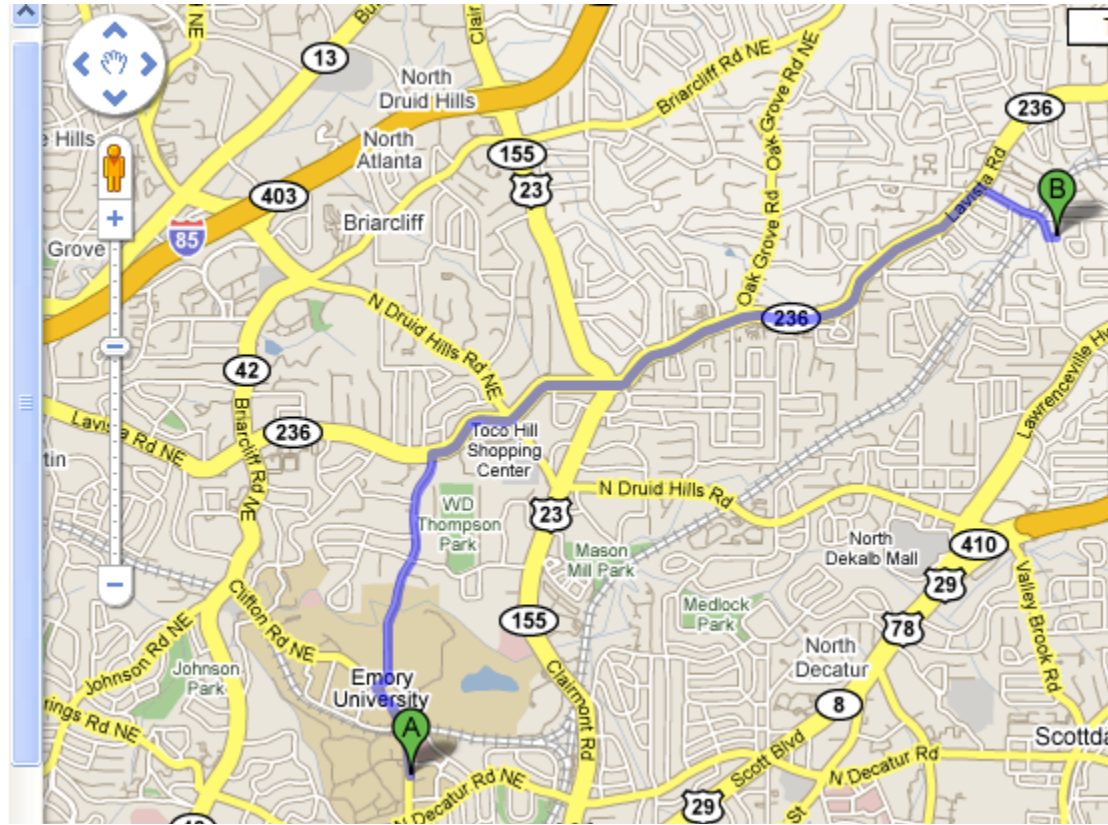
By car

Also available: [Public Transit](#) [Walking](#)

Driving directions to 3020 Frazier Woods Rd, Decatur, GA 30033
5.4 mi – about 12 mins

1364 Clifton Rd NE
Atlanta, GA 30322

1. Head north on Clifton Rd NE toward Woodruff Circle NE 0.5 mi
2. Turn right at Houston Mill Rd NE 1.2 mi
3. Turn right at GA-236/Lavista Rd NE 3.2 mi
4. Turn right at Frazier Rd 0.4 mi
5. Turn left at Frazier Woods Rd 200 ft



Research Relies on Accurate Data

- NAACCR Census Tract Certainty Codes

Code	Description
1	Census tract based on complete/valid street address
2	Census tract based on residence ZIP + 4
3	Census tract based on residence ZIP + 2
4	Census tract based on residence ZIP code only
5	Census tract based on ZIP code of P.O. Box
6	Census tract based on city or ZIP w/ one tract only
9	Unable to assign census tract

Research Relies on Accurate Data

- NAACCR GIS Coordinate Quality (abbreviated descriptions)

Code	Description
01	Coordinates assigned by Global Positioning System (GPS)
02	Coordinates are based on property parcel location
03	Coord are match interpolated over street segment's range
04	Coordinates are street intersections
05	Coordinates are at mid-point of street segment
06	Coordinates are address ZIP code+4 centroid
07	Coordinates are address ZIP code+2 centroid
08	Coordinates were obtained manually by lookup
09	Coordinates are address 5-digit ZIP code centroid
10	Coordinates are ZIP code of PO Box or Rural Route
11	Coordinates are centroid of address city
12	Coordinates are centroid of county


Example

Street Address Successfully Geocoded

3020 Frazier Wood Road Dec, GA 30033 EZ-Locate returned a segment match .		
Matched Address	Location (WGS-84)	2000 Census
3020 FRAZIER WOODS RD DECATUR, GA 30033	<u>Decimal Degrees</u> <u>Deg:Min:Sec</u> Lat: 33.830656 33:49:50.362N Lon: -84.265918 84:15:57.305W	FIPS State...: 13 FIPS County...: 089 Tract.....:0217.04 Block.....: 2011 FIPS Place...: FIPS MCD.....: 93096 MSA.....: 0520 CBSA.....: 12060 Metro.....: 0 Micro.....: 0
Standard Address	Postal	
3020 FRAZIER WOODS RD DECATUR,GA 30033-1843	Carrier.....: C014	

Source: U.S. Census Bureau, Census 2000 Summary File 3

Geography Identifier	Geography Identifier	Geog Summary Level	Geography	Individuals for whom poverty status is determined	Number; Below poverty level	Percent below poverty level
14000US13 089021704	13089021704	140	Census Tract 0217.04	5113	135	2.6

33.830656, -84.265918 

+33° 49' 50.36", -84° 15' 57.30"

Get directions: [To here](#) - [From here](#)
[Search nearby](#) - [Save to My Maps](#)

Frazier Rd

Frazier Woods Rd

Frazier Rd

Distance = 50.25 m

Frazier Woods Rd

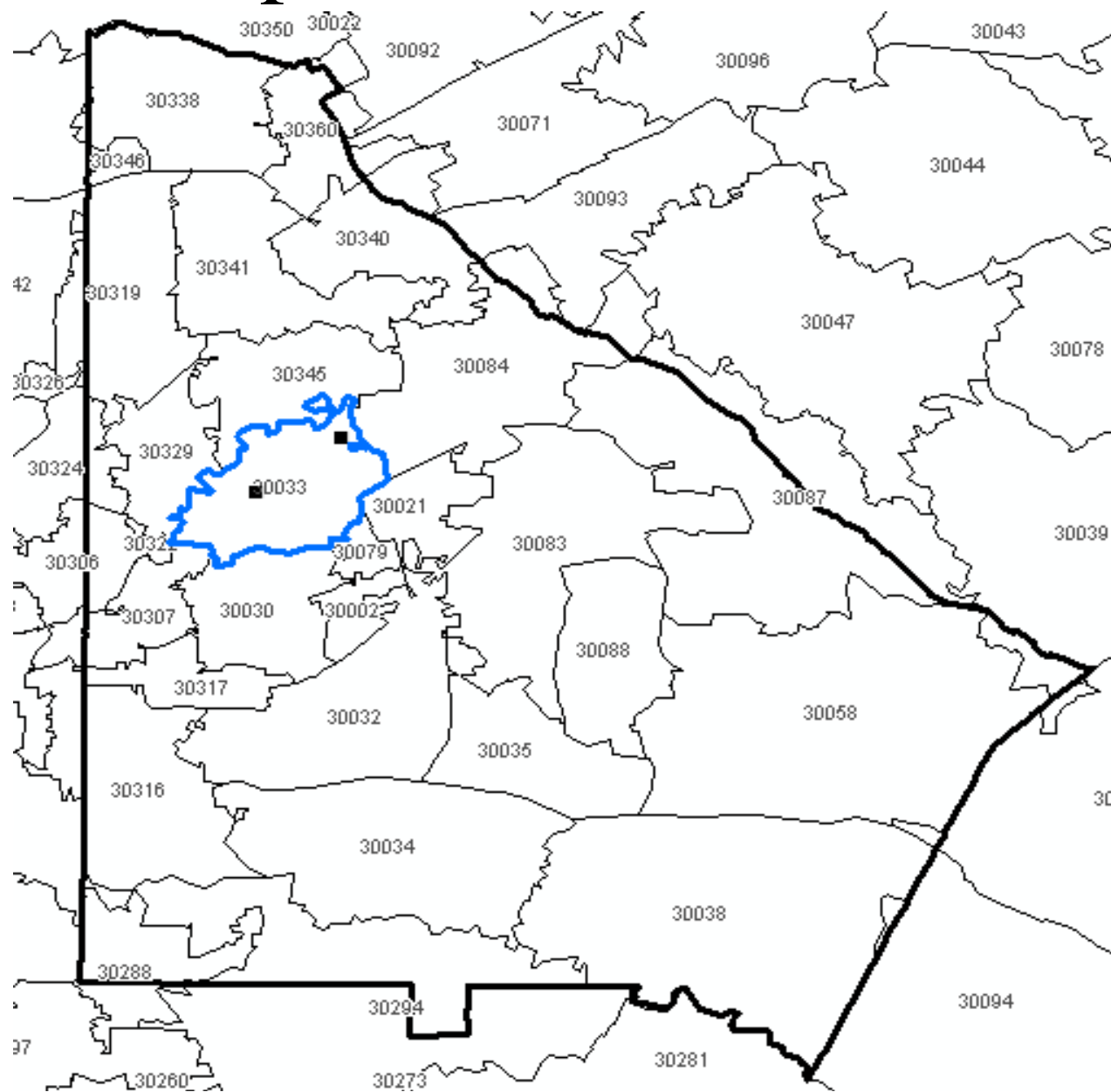


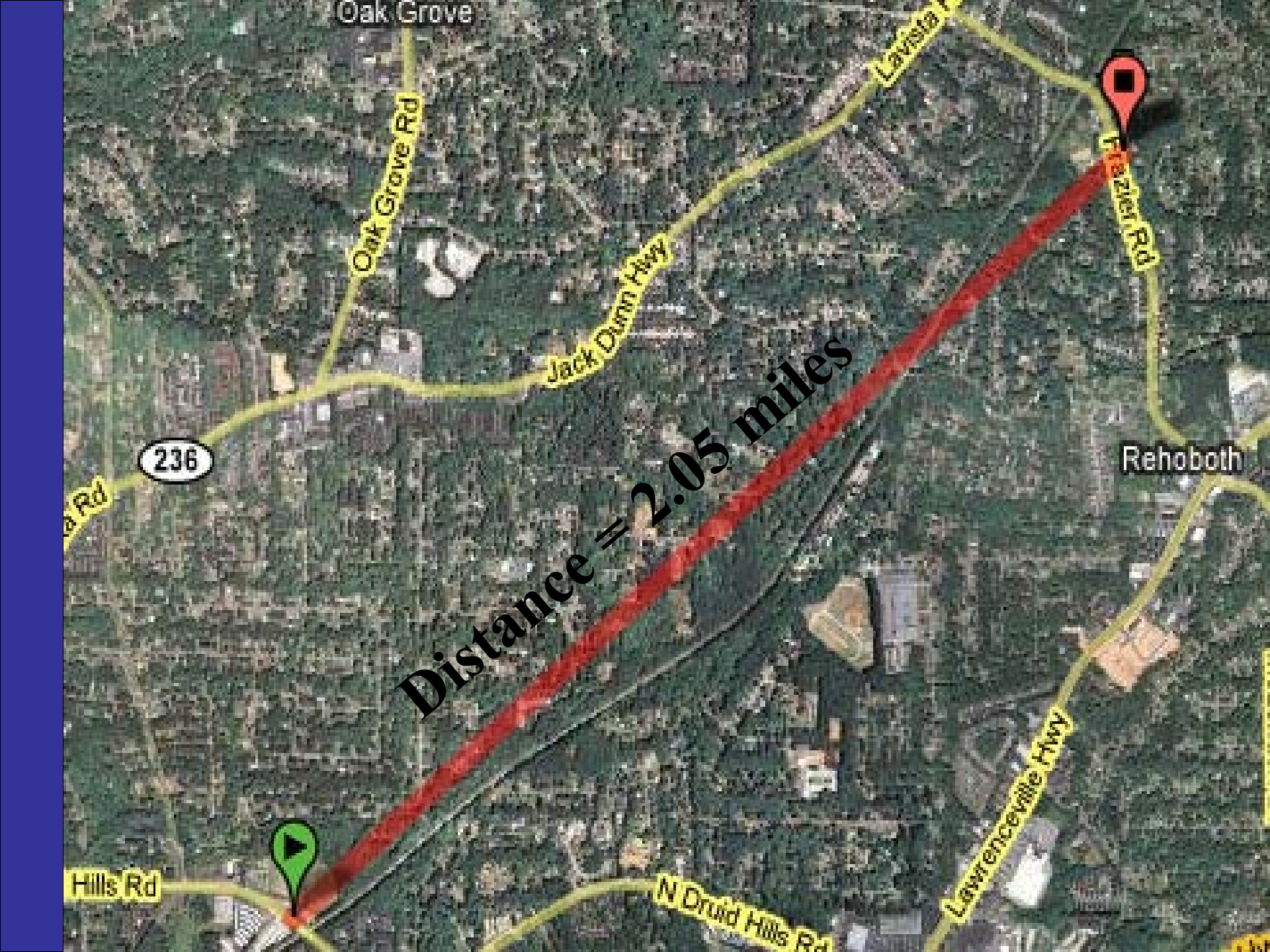
Example Continued

Error in street number causes a match to 5-digit zip code centroid

2030 Frazier Woods Road Decatur, GA 30033 EZ-Locate returned a 5-digit ZIP Code centroid.		
Matched Address	Location (WGS-84)	2000 Census
	<u>Decimal Degrees</u> <u>Deg:Min:Sec</u> Lat: 33.811830 33:48:42.588N Lon: -84.293717 84:17:37.381W	FIPS State...: 13 FIPS County...: 089 Tract.....:0216.02 Block.....: 2011 FIPS Place...: 56112 FIPS MCD.....: 90150 MSA.....: 0520 CBSA.....: 12060 Metro.....: 0 Micro.....: 0
Standard Address	Postal	
	Carrier.....:	

Compare Geocoded Points





Oak Grove

Lavista Hwy

Oak Grove Rd

Jack Dunn Hwy

Frazier Rd

236

Rehoboth

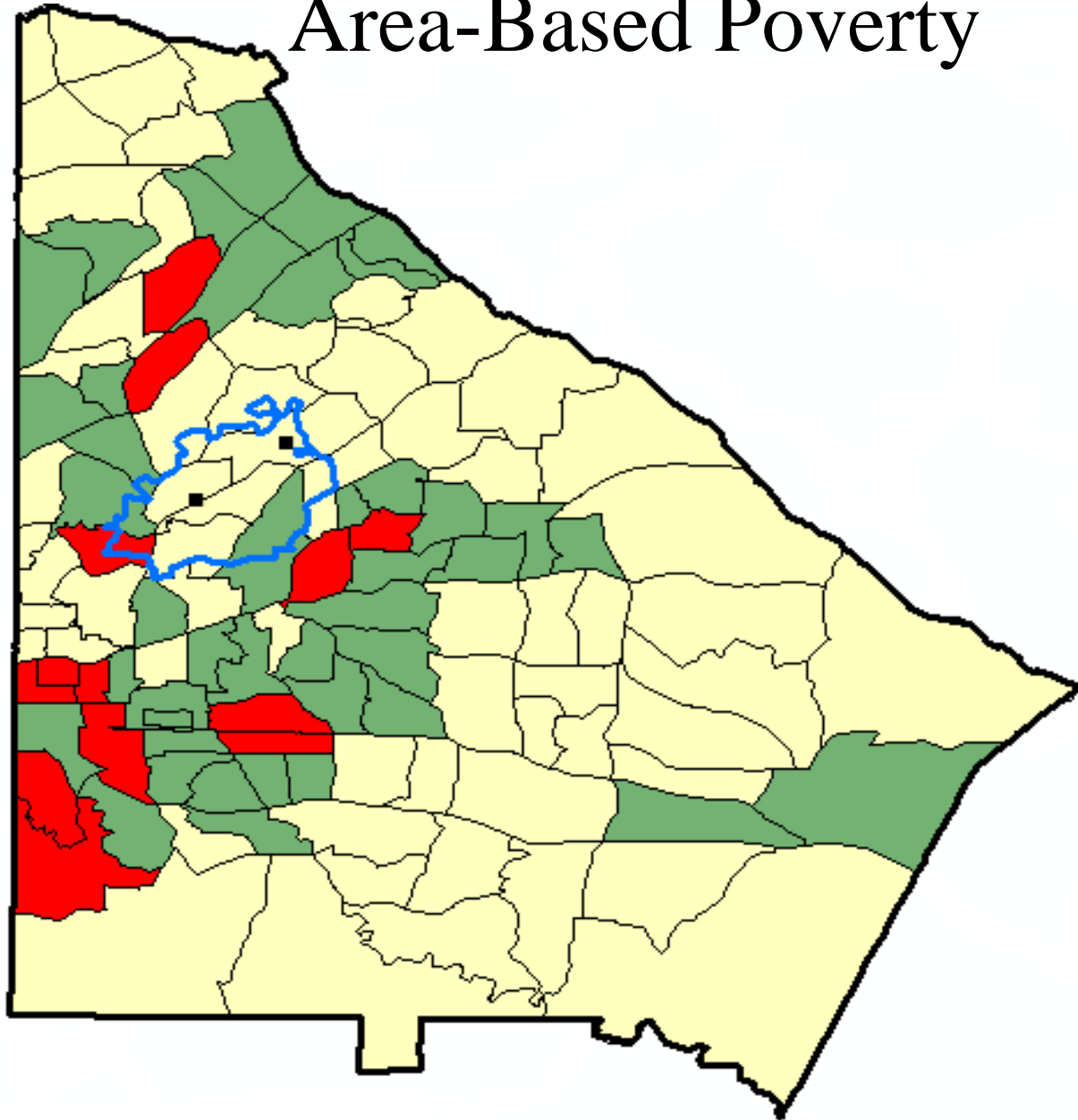
Distance = 2.05 miles

Hills Rd

Lawrenceville Hwy

N Druid Hills Rd

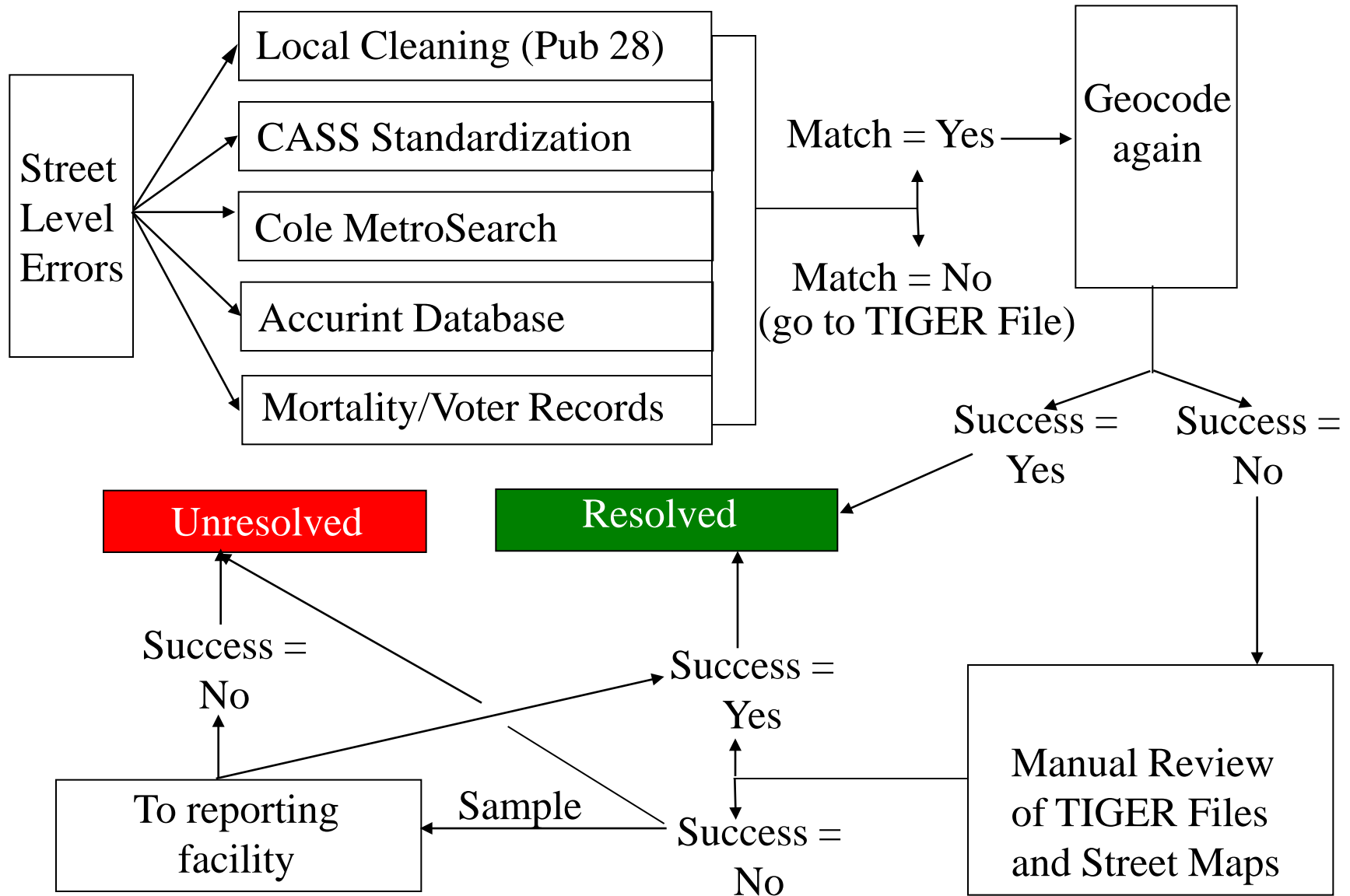
Compare Assignment of Area-Based Poverty



Small Geocoding Research Project

1996-2000 Data	Sample of GA Urban Counties	
	No.	%
<i>Total Records</i>	50,840	100.0%
PO Box	579	1.1%
Rural Route	7	<0.1%
Street Address	50,254	98.8%
Not Geocoded Certainty=1	4,486	8.8%

Flow Diagram of Steps to Clean Address Data



USPS Publication 28

Website: <http://pe.usps.gov/text/pub28/welcome.htm>

General tips for formatting address data

Example: The pound sign (#) should not be used as a secondary unit designator if the correct designation, such as APT or STE, is known. (100 Main ST APT 1)

If the pound sign (#) is used, there must be a space between the pound sign and the secondary number.

(100 Main ST # 1)

Address Standardization

- **CASS (Coding Accuracy Support System)** is a system the U.S. Postal Service uses to evaluate the accuracy of address-matching software programs.
- Address Standardization - Correct misspellings, directional, suffix and unit designator adjustments as directed by USPS CASS certification address correction standards. ZIP-code or city name address correction may be required. Append +4 to ZIP Codes.

Website CorrectAddress by Intelligent Search:

<http://www.intelligentsearch.com/address-verification/correct-address.html>

Examples of Standardization

- 1437 **MLK WY**, Atlanta, GA 30032
- 1437 **Martin Luther King Jr. Way**, Atlanta, GA 30032

- 800 **Lakridge** Dr **27**, Atlanta, GA 30032
- 800 **Lakeridge** Dr **STE 27**, Atlanta, GA 30032

- 400 **A** Peachtree **Av** NE, **Smyrna**, GA 30332
- 400 Peachtree **Ave** NE, **APT A**, **Atlanta**, GA 30332

Cole MetroSearch (Batch)

<http://www.coleinformation.com/>

MetroSearch for Windows Ver. 4.2 - [Reference Search]

File Edit Search Batch Dialer Queue Options Help

GA - GEORGIA

Last Name		Name: KEVIN C WARD					
First Name		Address: 3020 FRAZIER WOODS RD DECATUR, GA 30033-1843					
Middle Initial		Cross Street:					
House Number		Phone: (770) 939-6811					
Street Dir		County: DEKALB					
Street		Bus/Res: R					
Apt/Box #		Homeowner: Confirmed					
City		Residents			Gender	Date of Birth	Home Value:
State		KEVIN C			M		SIC Code:
Zip		HEATHER			F		Dwelling Units: 1
Phone	7709396811	Census Tract: 021704					
County		Length of Res: 9					

Household / Detail / Print 1 of 1

Click to select one Shift-Click to select a group Ctrl-Click to select multiples






Last Name	First Name	House Nur	Street Dir	Street	City	State	Zip
WARD	KEVIN	3020		FRAZIER WOODS RD	DECATUR	GA	30033

Number Selected: 1 Record Queue Count: 0

start e i f c m p w v i n b o x i n f o r m a t i o n ... O:\Appli... MetroS... Desktop 8:17 AM

Accurint (Batch)

<http://www.accurint.com>

All	Full Name	SSN	Address	Phone Information
 COPY	KEVIN C WARD Gender: Male	xxx-xx-xxxx	 3020 FRAZIER WOODS RD DECATUR GA 30033-1843 Jul 02 - Jul 08	770-939-6811 - EDT WARD KEVIN & HEATHER
We Also Found: <input type="checkbox"/> Property Records <input type="checkbox"/> Business Affiliations Purchase Results				
 COPY	KEVIN CHAD WARD Gender: Male	xxx-xx-xxxx	3111 WILTSHIRE DR AVONDALE ESTATES GA 30002-1439 Sep 98 - Jun 02	404-294-1739 - EDT
 COPY	KEVIN CHAD WARD Gender: Male	xxx-xx-xxxx	7312 NE CLAIRMONT RD ATLANTA GA 30329 Oct 97	404-321-1235 - EDT
 COPY	KEVIN CHAD WARD Gender: Male	xxx-xx-xxxx	5107 ASHFORD GABLES DR ATLANTA GA 30338-6780 Dec 94	294-1739

Mortality and Voter Records

- Mortality
 - only use if address at death matches address at diagnosis but provides more complete information (or death closely follows diagnosis)
- Voter
 - Voter files do not allow PO Box for residence address
 - Need to verify that address was the same both before and after the cancer diagnosis

Small Geocoding Research Project

1996-2000 Data	Sample of GA Urban Counties	
	No.	%
<i>Records Cleaned/Geocoded</i>	4,076	90.9%
PO Box	481	83.1%
Rural Route	5	71.4%
Street Address	3,590	92.1%

Results of Data Clean-up by Source

Additional Source	Existing Source					
	Accurint	Cole	CASS	Local	Voter	Mortality
Accurint	75.10%	58.70%	38.10%	53.70%	39.80%	62.60%
Cole	3.20%	19.60%	10.10%	14.40%	7.20%	17.50%
CASS	17.60%	45.10%	54.60%	22.20%	33.10%	46.60%
Local	11.10%	27.30%	0.00%	32.50%	19.50%	27.00%
Voter	6.60%	29.50%	20.40%	29.00%	41.90%	35.30%
Mortality	5.60%	16.00%	10.10%	12.70%	11.50%	18.10%

Evaluation of Misclassification by Poverty

	<u>Residence ZIP Centroid</u>	<u>PO ZIP Centroid</u>
<hr/> Misclassification by: <hr/>		
Tract		
Percent	59.5%	81.8%
Confidence Interval	(57.9, 61.2)	(78.0, 85.0)
Tract Poverty 2-groups*		
Percent	8.0%	18.9%
Confidence Interval	(7.2, 9.0)	(15.6, 22.6)
Tract Poverty 3-groups [#]		
Percent	20.9%	43.8%
Confidence Interval	(19.6, 22.3)	(39.4, 48.3)

[#] Census assigned poverty [% living below poverty line]: (0-9.9, 10-19.9, 20+)

* Census assigned poverty [% living below poverty line]: (0-19.9, 20+)

Take Home Points

- Review geocoding certainty variables in your own data to understand the quality of the data and areas for improvement.
- When geocoded Registry data is used for research, **ALWAYS** provide certainty variables to researchers.
- At a minimum, standardize your data prior to geocoding. Accurint is a nice source for cleaning older data but requires some resources and effort.


Geocoding Best Practices

- www.NAACCR.org

**A Geocoding
Best Practices Guide**

By Daniel W. Goldberg
University of Southern California
GIS Research Laboratory

November 2008



SPONSORING ORGANIZATIONS:
Canadian Association of Provincial Cancer Agencies
Canadian Partnership Against Cancer
Centers for Disease Control and Prevention
College of American Pathologists
National Cancer Institute
National Cancer Registrars Association
Public Health Agency of Canada

SPONSORS WITH DISTINCTION:
American Cancer Society
American College of Surgeons
American Joint Committee on Cancer

NAACCR North American Association of Central Cancer Registries, Inc.

An aerial topographic map of the United States, showing state boundaries and terrain features. The map is centered on the continental United States, with the Great Lakes region to the north and the Gulf of Mexico to the south. The text "Thank You." is overlaid in the center of the map.

Thank You.

Questions?