

Obtaining Disparities-Relevant Information from the US Census



Doug Einstadter, MD, MPH

Better Health Greater Cleveland

September 12, 2011

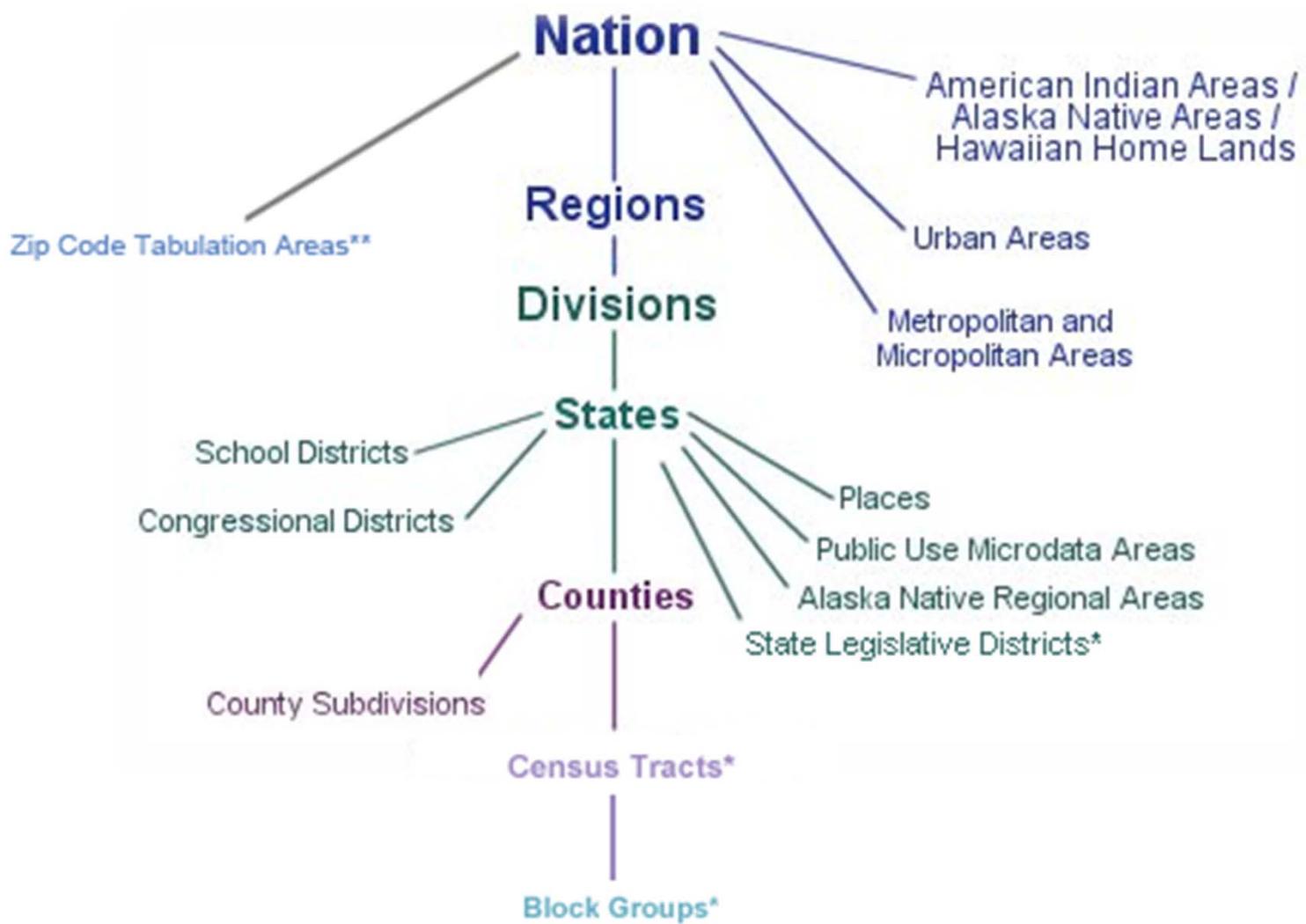
Outline

- **Overview of Census Geography.**
- **Process for geocoding patients' addresses.**
- **The census data we use characterize vulnerable populations.**
- **What we provide to our partners to link their geocoded data to the census.**

Outline (cont.)

- The fuzzy factor: HIPAA compliance and maintaining patient confidentiality.
- Demonstration of Excel application.
- Thoughts about using the ACS.

US Census Geographic Hierarchy



Census Geography

Blocks

A **census block** is a component of a block group. Blocks are small in area, in general, especially in cities. However, blocks in rural or remote areas may cover hundreds of square miles. There are 11,078,297 blocks in the Census 2010 geography.

Census Geography

Block Groups (BGs)

A **block group** is a collection of one or more census blocks and a statistical division of a census tract. Block groups do not cross census tract, county, or state boundaries. In general, a block group is comprised of 600 to 3,000 residents. There are 217,740 block groups in the Census 2010 geography.

Census Geography

Census tracts

Census tracts are small statistical subdivisions of a county, typically with 1,200 to 8,000 residents. The boundaries are usually delineated by local committees, and do not cross county or state lines. There are 73,057 tracts in the Census 2010 geography.

Census Geography

Residential ZIP Codes

Created by the U.S. Postal Service to deliver the mail, ZIP Codes *do not represent standard census geographic areas for data reporting*. Because ZIP Code boundaries are not contiguous with census geographic areas or stable over time, data estimated for ZIP Codes are also subject to change. Residential ZIP Code data are estimated from census block group data.

The American Community Survey

- The American Community Survey (ACS) is a household survey conducted by the U.S. Census Bureau with an annual sample size of about 3 million addresses.
- The ACS Summary file Includes all the estimates and margins of error from the Detailed Tables and geographies that are published for the ACS.
- Has about 350 tables with data at the Census Block Group level.

The American Community Survey

- Contains over 18 billion cells; the tables are stored in a series of files (n=12,5000 at the census block level) organized by state and ‘sequence number.’
- The files are divided into three types:
 - Geographies – (position based)
 - Estimates – (comma delimited)
 - Margins of Error – (comma delimited)
- Require the user to merge the file types to obtain final data file.

The American Community Survey

- Data contained in the ACS Summary File cover demographic, social, economic, and housing subject areas.
- Only the 5-year data includes data to the census tract and block group level.
- For the 2005-2009 ACS 5-year data release, data for census block groups are only available using the ACS Summary File.

Geocoding



Geocoding

- The process of finding associated geographic coordinates (often expressed as latitude and longitude) from other geographic data, such as street addresses, or zip codes (postal codes).
- A simple (relatively) method for applying neighborhood attributes to individuals
- A geocoder is a piece of software or a (web) service that helps in this process.

Geocoding

Requirements:

- Geocoding software or service.**
- Clean patient addresses (street, city, zip) for Block level data; zip code for ZCTA level.**
- Census data**
 - 2000 US Census Summary File 3.**
 - American Community Survey (ACS) 5-year (2005-2009) Summary file estimate.**

Examples of Available Area-Based Socioeconomic Measures

- ***Educational level - % HS graduation**
- ***Median household income (\$)**
- **Percent of residents below poverty**
- **Percent English speakers**
- **Percent foreign born**
- **Percent female head of household**
- **% Employed**

*Used by Better Health Greater Cleveland

Data Preparation Process

Data Preparation Process

- 1. Geocode patient addresses;**
- 2. Use “spatial join” to obtain census block (or tract) for each individual;**
- 3. Apply census attributes to each individual whose address lies inside the census block;**
- 4. ‘Fuzz’ the results to preserve anonymity;**

Address Clean-up

Addresses need to be standardized and have extraneous information removed. PO Box addresses are excluded.

Examples:

12555 Bellaire Rd Apt 101 -----> 12555 BELLAIRE RD

5500 Northfield RD #522 -----> 5500 NORTHFIELD RD

3068 W 47 ST Down -----> 3068 W 47 ST

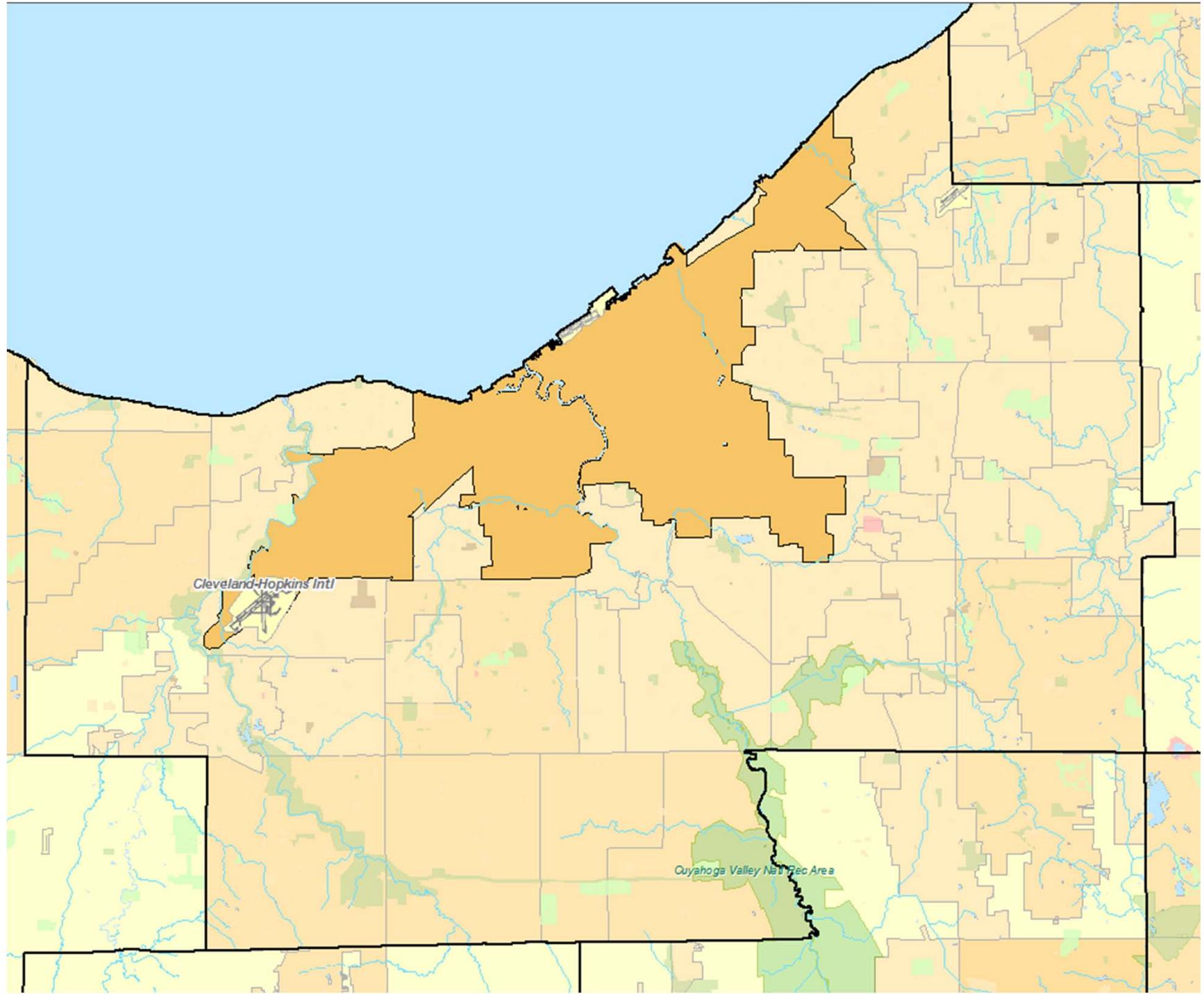
3218 E 117 ST UP -----> 3218 E 117 ST

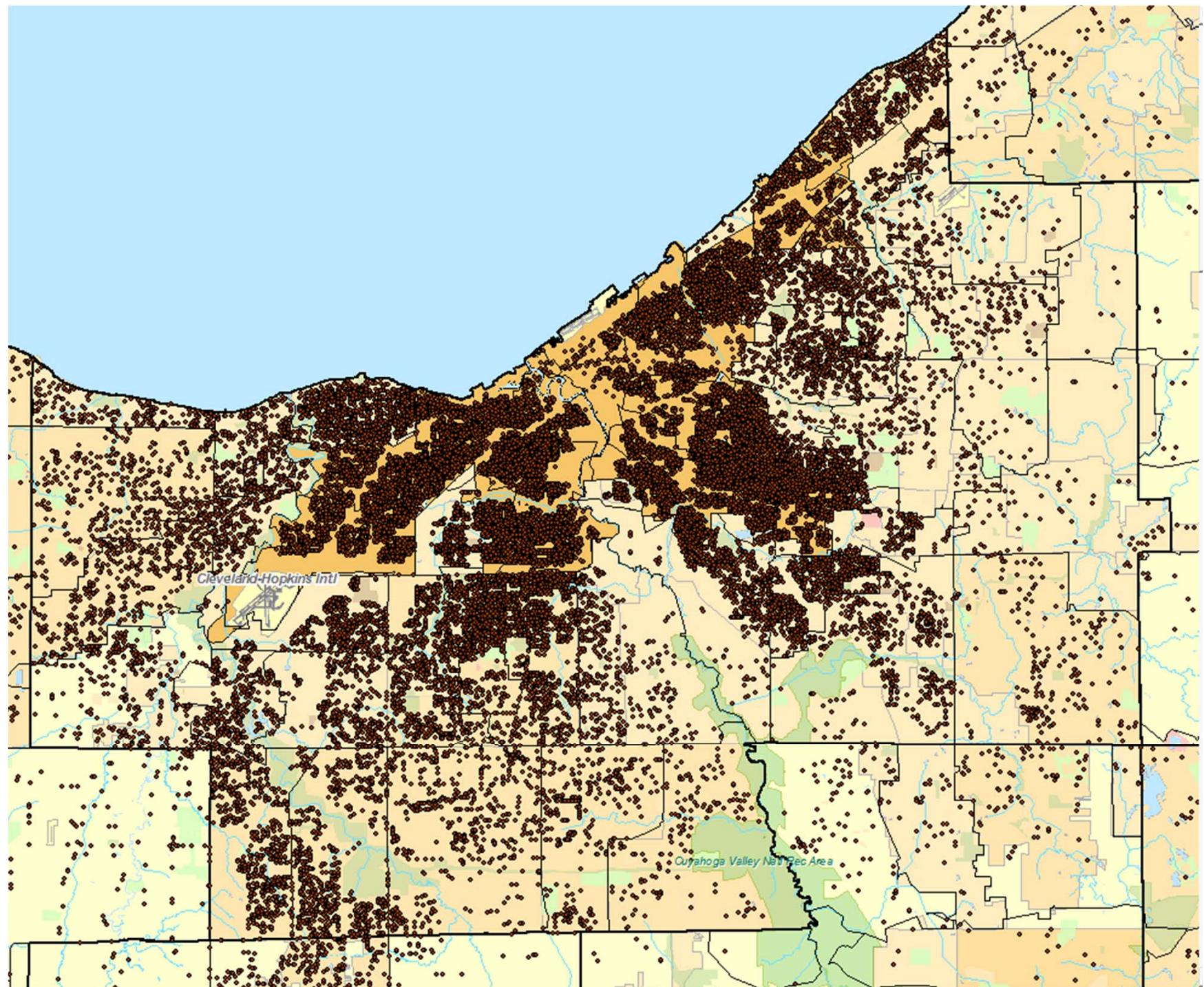
1500 E 193 ST APT 322 BUILDING E -----> 1500 E 193 ST

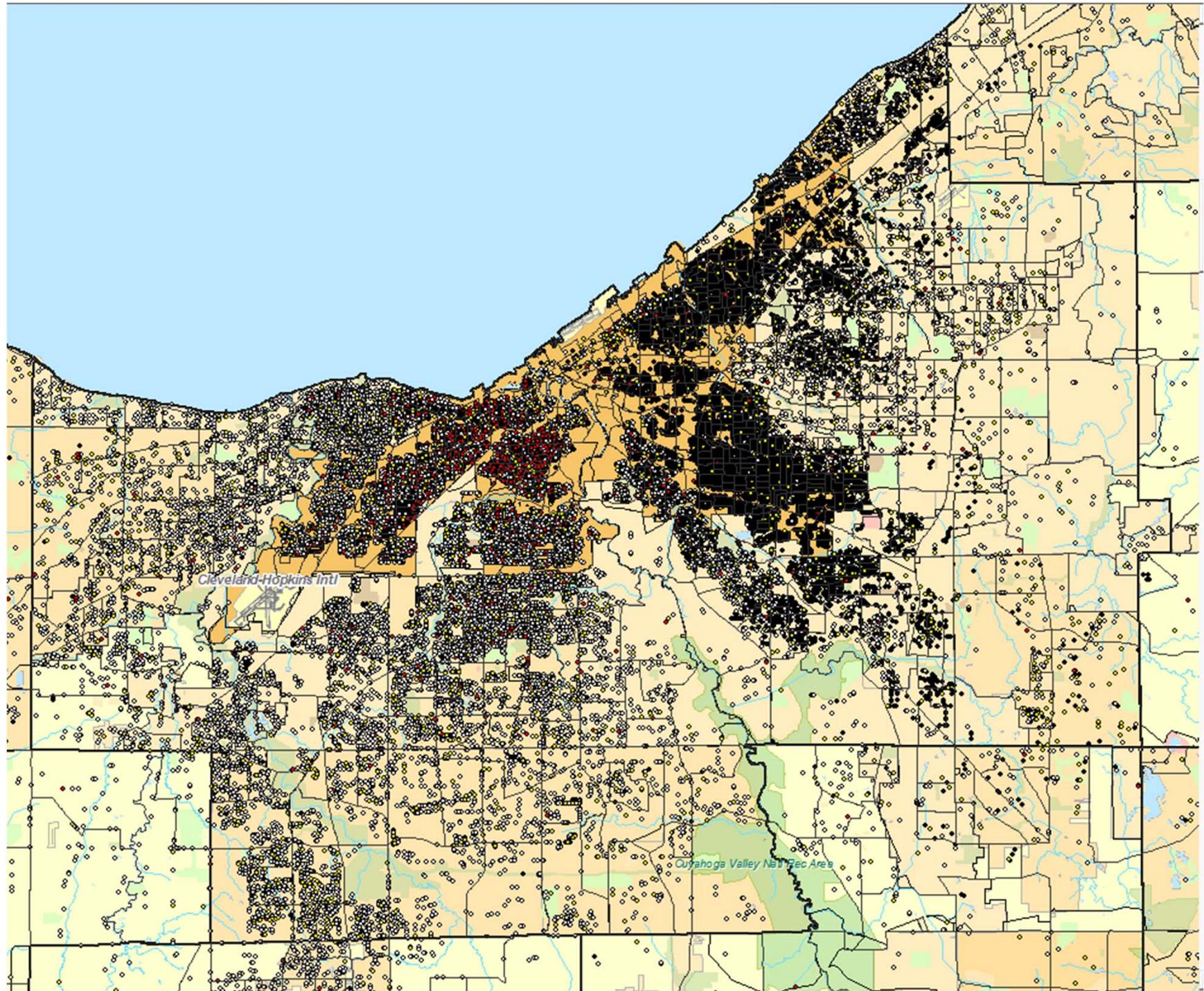
1654 WARREN Rd Unit Down -----> 1654 WARREN RD

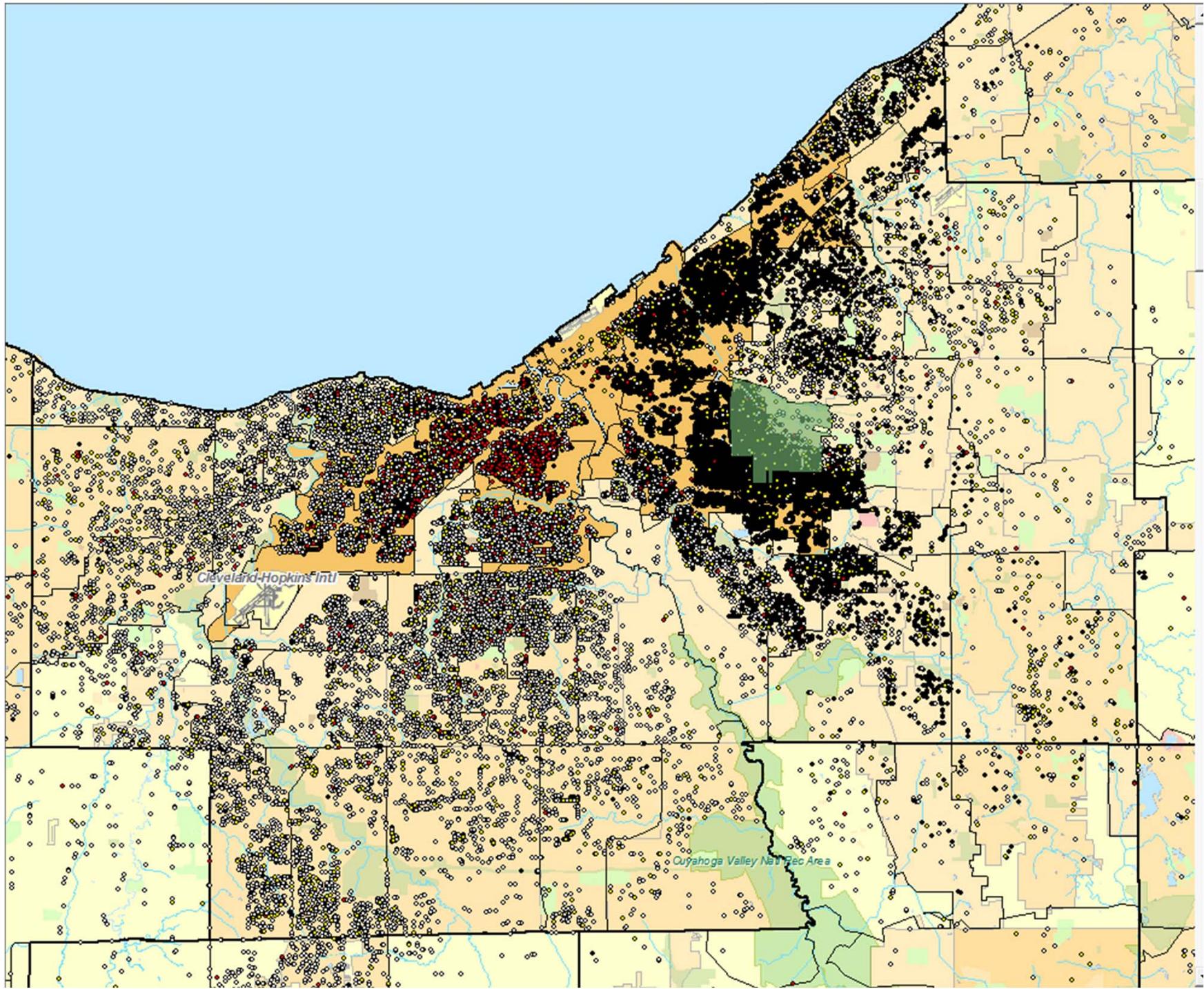
4705 HARVARD AVE down front -----> 4705 HARVARD AVE

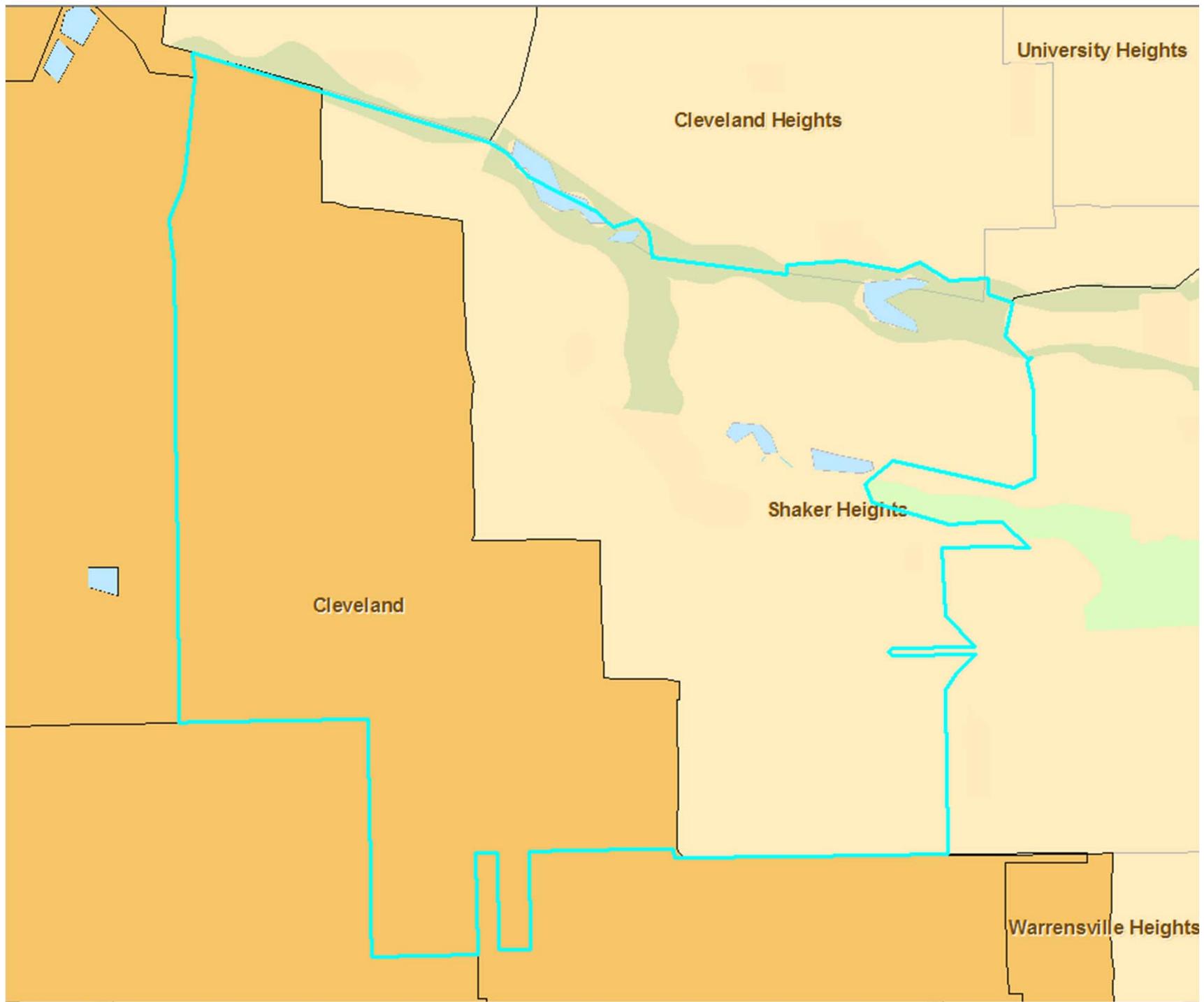
Mapping the Data

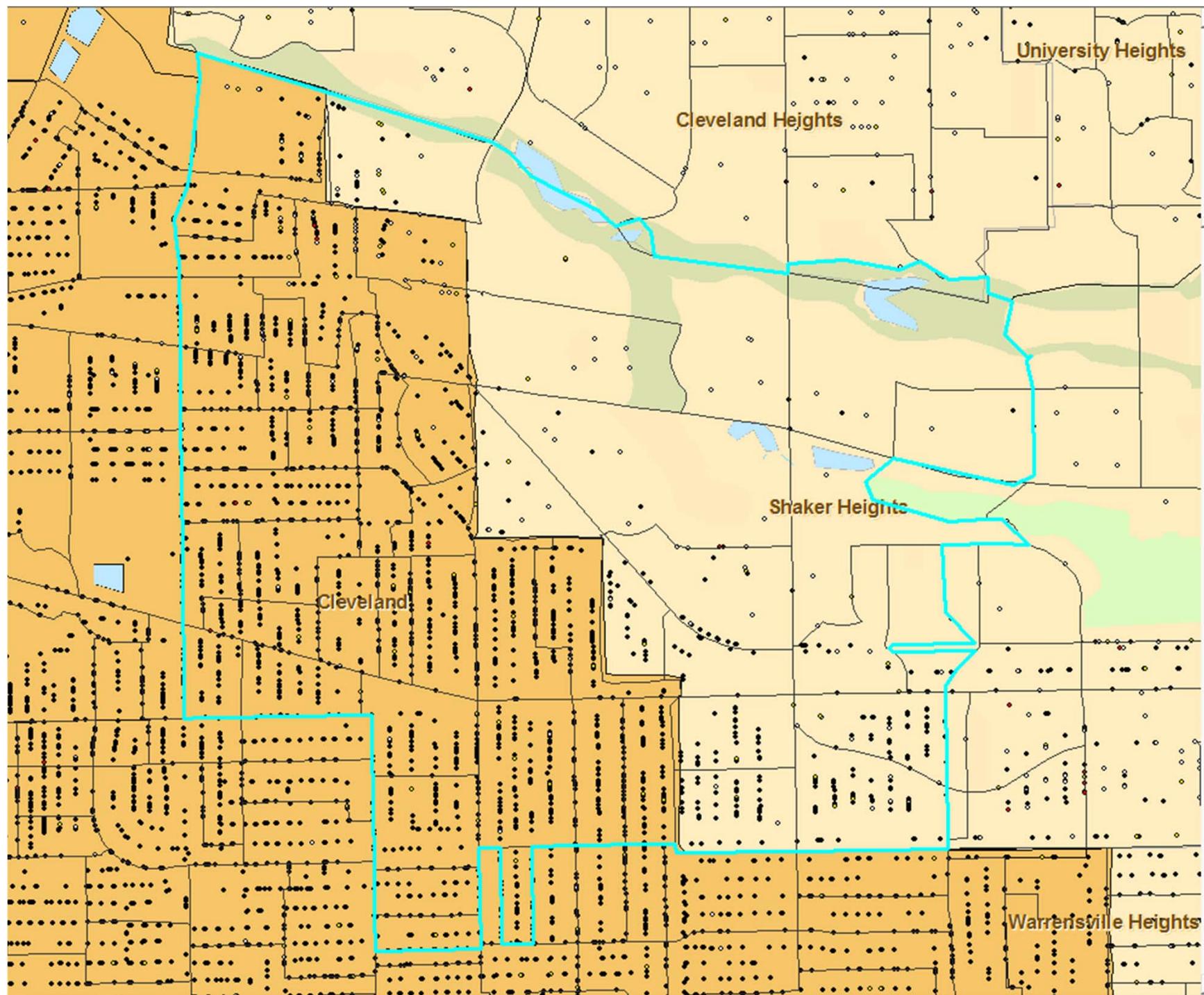


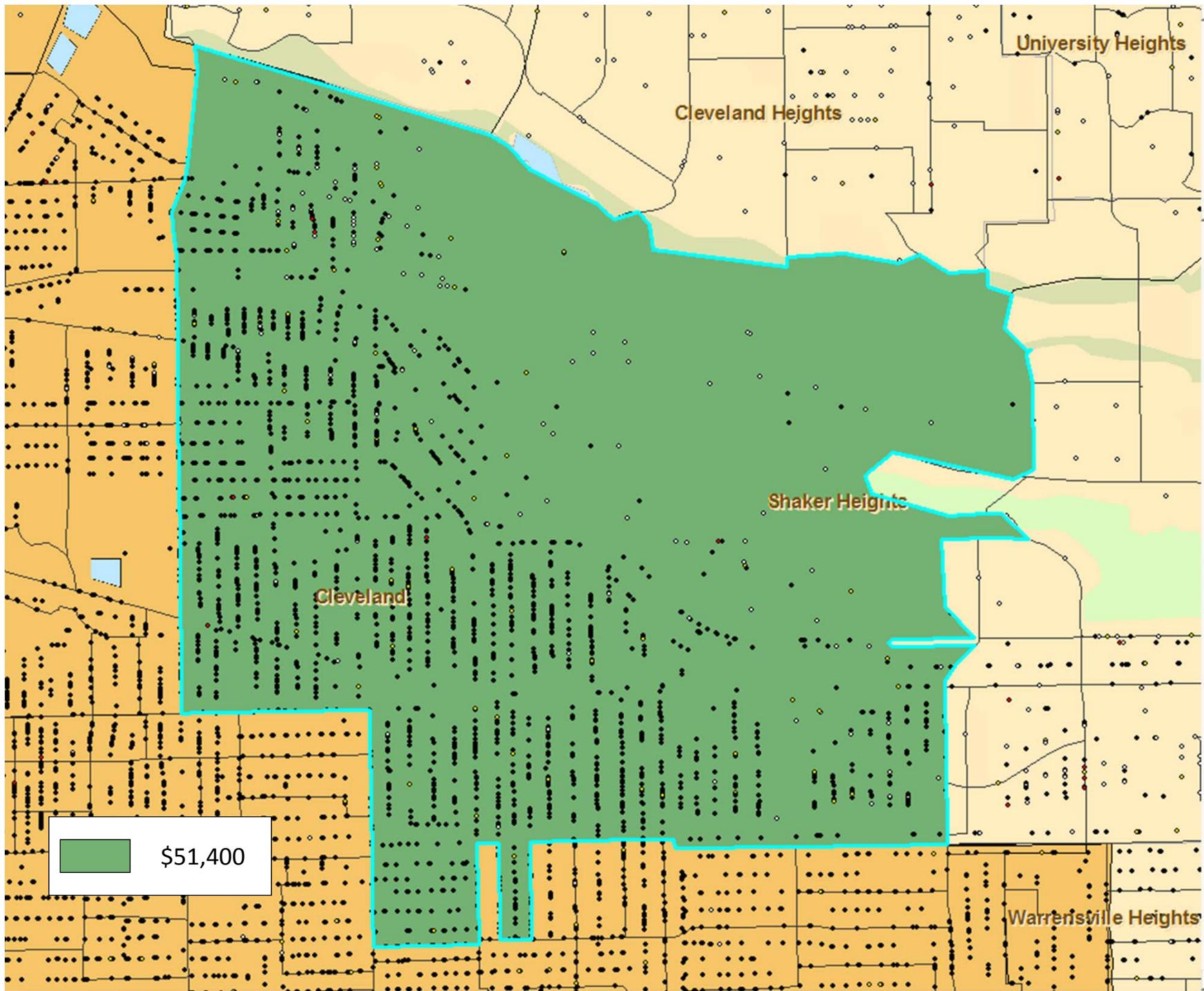


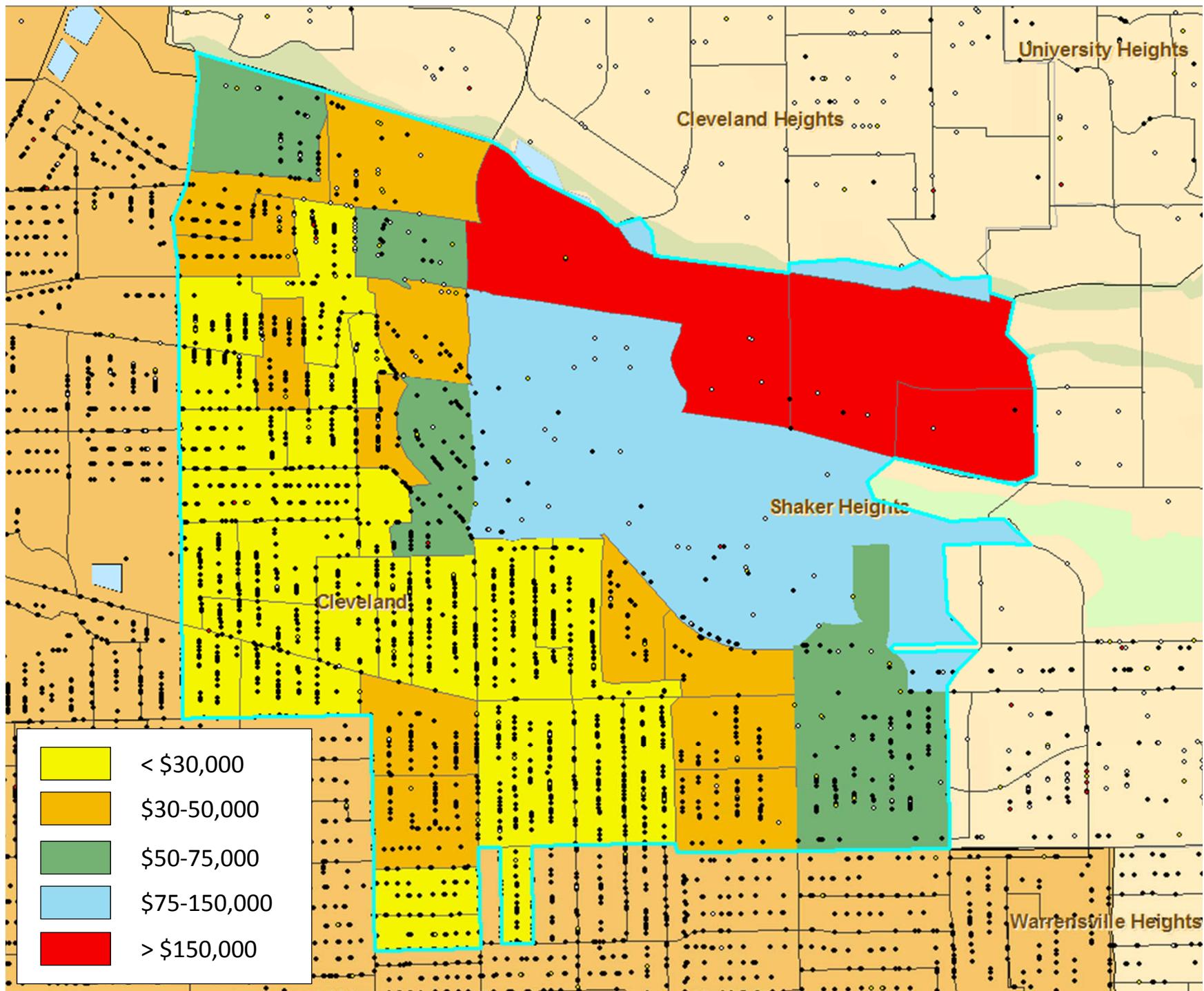












Excel File

**Adding some randomness
to the results**

How We Have Used Census-Based Estimates in Reporting

Better Health *Greater Cleveland*

Thomas.Love@case.edu

September 12, 2011

www.betterhealthcleveland.org

Thomas E. Love, Ph.D.

Associate Professor, Medicine, Epidemiology & Biostatistics, CWRU

Director, Data Management Center, *Better Health*

Co-Chair, Information Management Committee, *Better Health*



An Alliance for Improved Health Care

A project of Better Health Greater Cleveland,
the Robert Wood Johnson Foundation's Aligning
Forces for Quality initiative and other funders.

Some Relevant Data Elements

CORE DATA ELEMENTS TO BE GATHERED ON ALL REPORTED PATIENTS (regardless of chronic condition)				
REPORTING PERIOD for Report 8: start date: July 1, 2010 through end date: June 30, 2011				
#	Variable	Name	Description / Codes	Comments / Business Rules
C-14	Patient's (Primary) Insurance Class	INSUR	0 = Missing or Unavailable 1 = Commercial or Private 2 = Medicaid 4 = Medicare 8 = Veterans 16 = Uninsured or Self-Pay 32 = Other (use sparingly)	Do not double code, but instead select a single primary class for each patient. For patients with multiple primary classes, isolate to 1, 2, 4 or 16 whenever possible. BHGC hierarchy for this purpose is Commercial, then Medicaid, then Medicare. 0 = Missing (includes not documented, unknown, unrecorded, etc.) 1 = Commercial or Private (includes "Medicare + Private") 2 = Medicaid (includes "Public NonCHIP, Medicaid/Medicare")
C-15	Patient's Preferred Spoken Language for Receipt of Medical Care	LANG	0 = Missing or Unavailable 1 = English 2 = Spanish 4 = Other Language (and see LANGTEXT)	0 = Missing (includes not documented, unknown, unrecorded, etc.) 1 = English 2 = Spanish 4 = Other Language (Specific Language to be reported in LANGTEXT)
C-16	Patient's Preferred Spoken Language for Receipt of Medical Care (Text specifying "Other" LANG)	LANGTEXT	Text to indicate preferred language of patient; i.e. Russian, French, Tagalog, etc. as available in medical record.	Patients with LANG = 4 must be reported here. Sites may also list text information for English and Spanish speaking patients, although this is optional.
C-17	Geocoding Method	GEOMETH	0 = Missing or Not Performed 1 = Zip Code 2 = Census Tract 4 = Census Block	All patients must be geocoded to obtain NINCOME, NEDLEV, CUYA and CLEVE information. Please indicate the method used - zip code is the least granular, census block the most.
C-18	Neighborhood Income Level based on Geocoding	NINCOME	Neighborhood Income Level (median income of home neighborhood) as estimated from Census 2000 via geo-coded patient address.	Report in \$, with no decimal places. "Fuzzed" results for zip code based geocoding are available through our February 2008 BHGC software.
C-19	Neighborhood Educational Attainment Level based on Geocoding	NEDLEV	Neighborhood Educational Attainment Level (% High School Graduate or more) estimated from Census 2000 via geo-coded address.	Report as a percentage, with two decimal places (% with education = HS graduate or more.) "Fuzzed" results for zip code based geocoding are available through our February 2008 BHGC software.
C-12	Patient's Race/Ethnicity (Gold Standard is Patient Self-Identification)	RACE	0 = Missing or Unavailable 1 = White or Caucasian 2 = African-American or Black 4 = Hispanic 8 = Other	0 = Missing (includes not documented, unknown, unrecorded, etc.) 1 = White or Caucasian 2 = African-American or Black 4 = Hispanic 8 = Other (includes Asian, Native Hawaiian, Pacific Islander, American Indian, Alaska Native and More than One Race)

2010 Community Health Checkup

TABLE 1. CHARACTERISTICS OF PATIENTS INCLUDED IN THIS REPORT

	Diabetes		High Blood Pressure		Heart Failure	
# of Patients	28,997		108,608		5,251	
# of Primary Care Practices	48 (8 health systems)		48 (8 health systems)		33 (3 health systems)	
	Better Health Population	Range of Values Across Sites	Better Health Population	Range of Values Across Sites	Better Health Population	Range of Values Across Sites
Insurance (%)						
Medicare	35.0	0 – 48	43.2	0 – 61	72.5	18 – 85
Commercial	43.3	0 – 74	41.4	0 – 78	19.2	2 – 40
Medicaid	8.9	0 – 39	6.3	0 – 37	5.2	0 – 34
Uninsured	12.8	0 – 100	9.1	0 – 100	3.1	0 – 21
Medicaid + Uninsured	21.7	0 – 100	15.4	0 – 100	8.3	0 – 49
Race/Ethnicity (%)						
White	52.6	2 – 96	60.8	2 – 98	64.6	3 – 97
African-American	39.6	1 – 97	34.5	0 – 97	32.0	0 – 97
Hispanic	4.6	0 – 64	2.2	0 – 54	1.9	0 – 46
Other	3.2	1 – 64	2.5	0 – 52	1.5	0 – 27
Non-White	47.4	4 – 98	39.2	2 – 98	35.4	3 – 97
Preferred Language (%)						
English	95.9	35 – 100	97.1	42 – 100	96.2	53 – 100
Spanish	2.2	0 – 57	1.1	0 – 51	1.2	0 – 48
Other Languages	1.9	0 – 63	1.8	0 – 57	2.6	0 – 30
Average Age	57.7	50 – 62	62.0	50 – 69	70.7	57 – 76
% Female	53.7	35 – 75	57.4	32 – 79	50.2	27 – 70
Median Household Income (\$)	41,200	25,500 – 68,000	44,300	25,300 – 71,200	43,100	25,000 – 69,000
High School Graduation Rate (%)	79.6	64 – 90	81.7	66 – 92	80.9	65 – 91
Average Body Mass Index	34.1	29 – 36	31.7	28 – 35	Not reported.	
% Not Smoking	79.7	42 – 92	82.0	31 – 92		

Diabetes Standards

TABLE 3. BETTER HEALTH'S INDIVIDUAL AND COMPOSITE STANDARDS FOR DIABETES

CARE PROCESSES

4 standards for good routine care:

- Blood sugar control test done
- Screening or treating kidney problems
- Annual eye exam
- Pneumonia vaccine given

Composite Reported: Percentage of patients who met all 4 standards

CLINICAL OUTCOMES

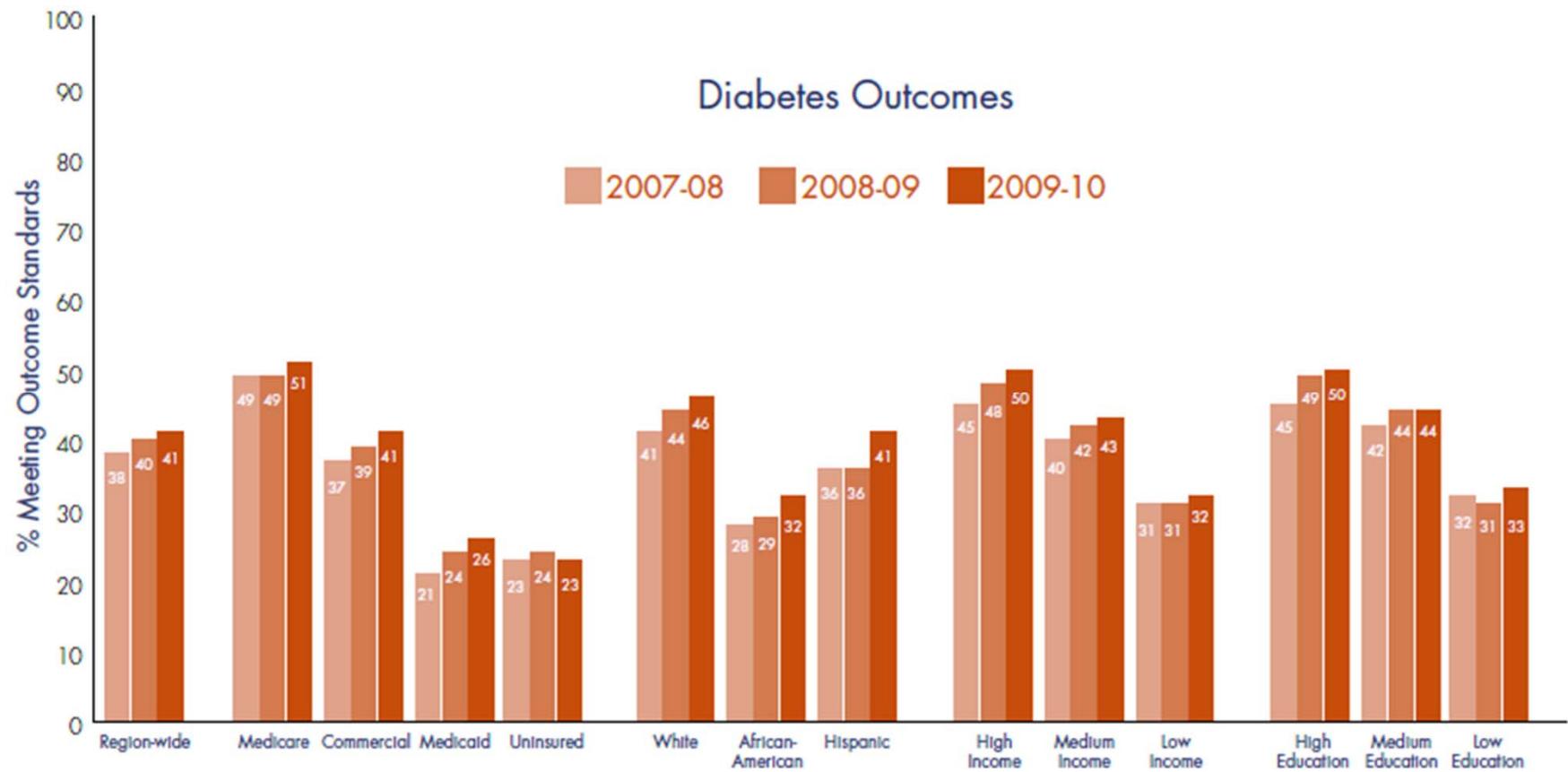
5 standards of good control:

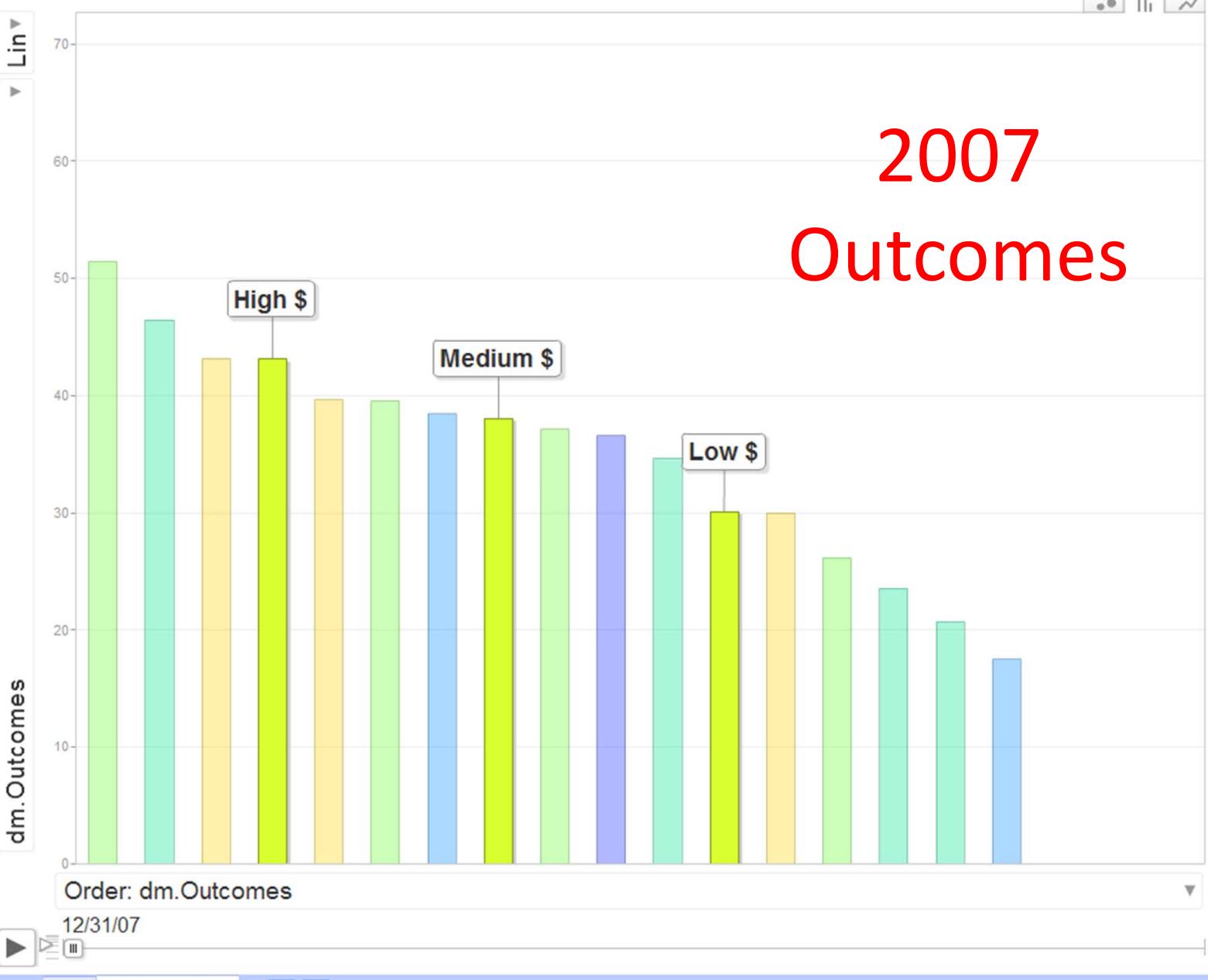
- Blood Sugar ($\text{HbA1c} < 8\%$)
- Blood Pressure ($< 140/80$)
- Cholesterol ($\text{LDL Cholesterol} < 100$ or statin)
- Weight (Body Mass Index < 30)
- Documented non-smoker

Composite Reported: Percentage of patients who met at least 4 standards

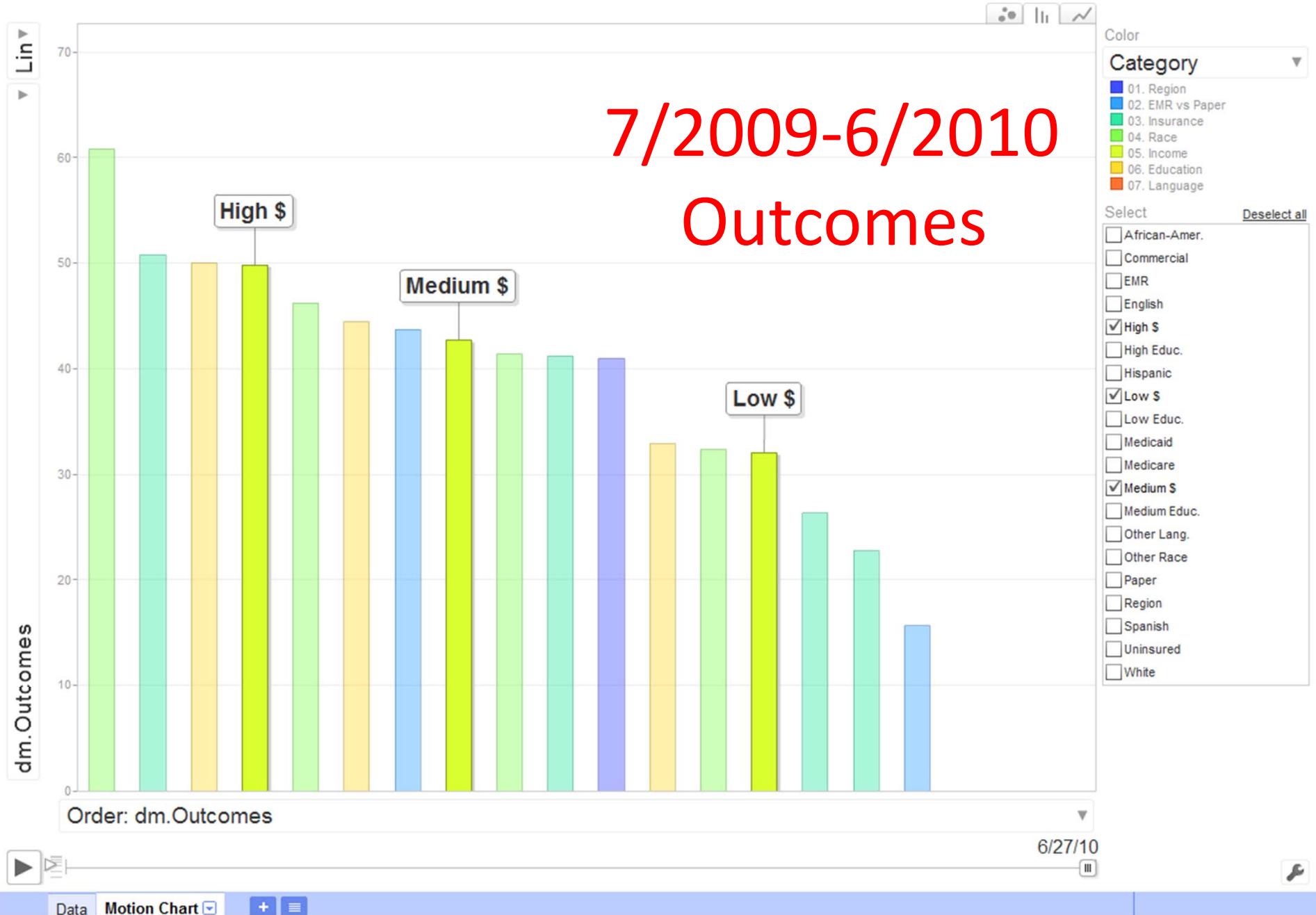


Tracking Change Over Time





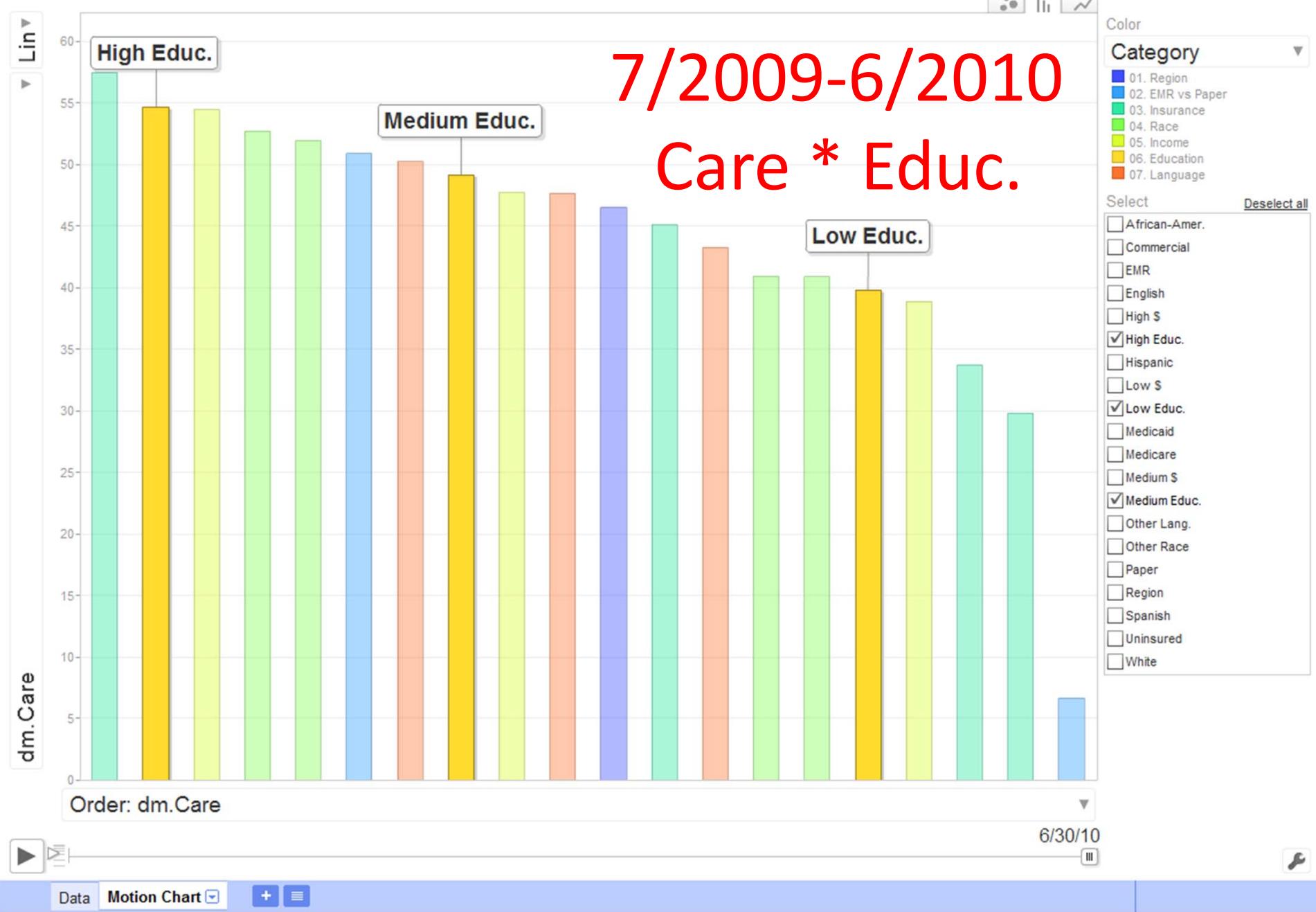
7/2009-6/2010 Outcomes





7/2009-6/2010

Care * Educ.



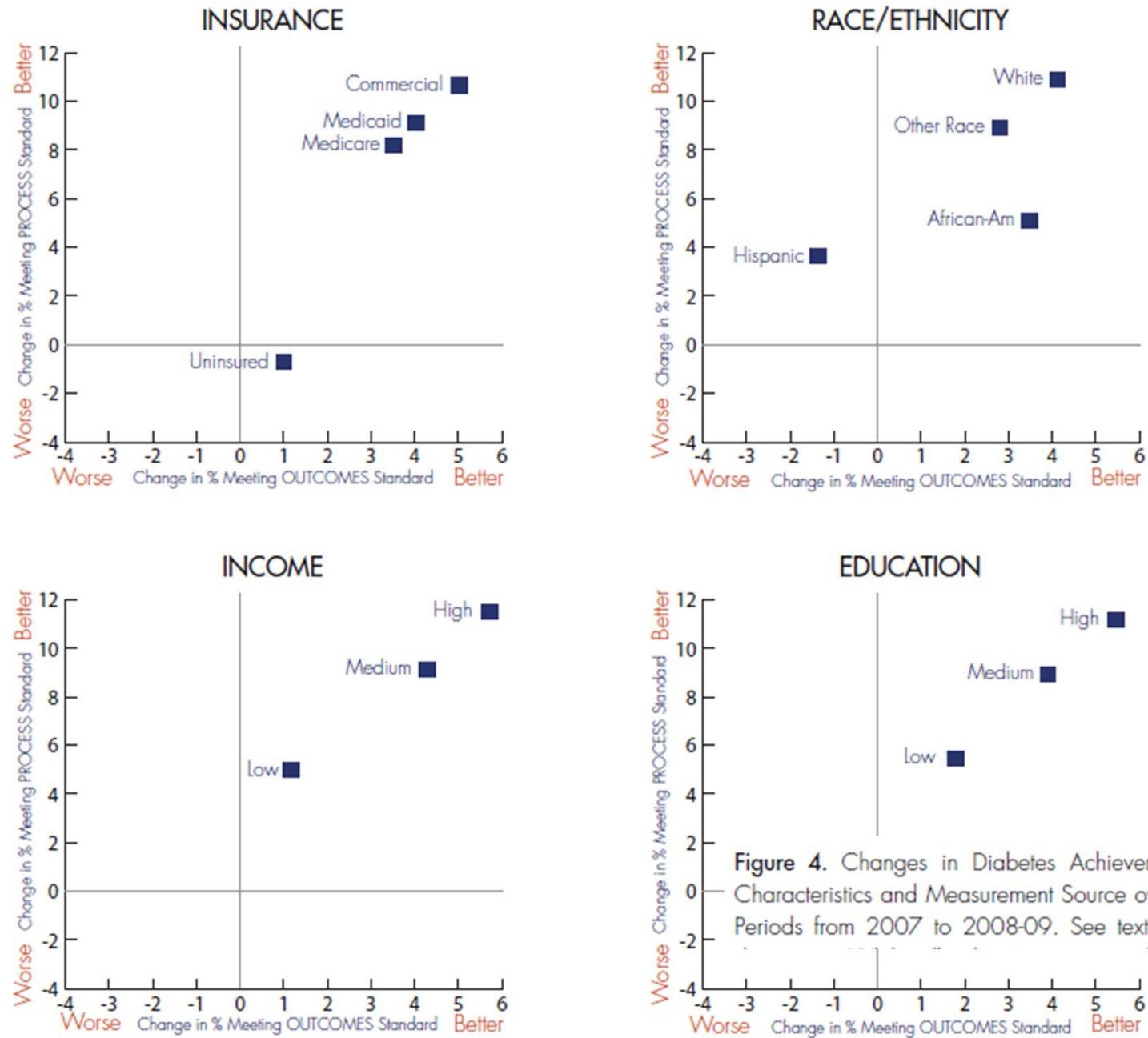


Figure 4. Changes in Diabetes Achievement by Patient Characteristics and Measurement Source over all Reporting Periods from 2007 to 2008-09. See text on Page 6 for

TABLE 4. BETTER HEALTH'S INDIVIDUAL AND COMPOSITE STANDARDS FOR HEART FAILURE

Evaluation Standards 4 Standards of Good Assessment	Treatment Standards 2 Types of Evidence-Based Medications
Heart Function Test Done ("Echo" to see how well your heart is pumping) Blood Test Done Each Year (Basic Metabolic Panel to check blood chemistry) Weight Checked Regularly (Look for fluid retention to monitor heart function) Blood Pressure Checked Regularly (High Blood Pressure can signal serious heart problems)	ACE/ARB Medication (Improves heart and kidney function and lowers blood pressure) Beta-Blocker Treatment (Blocks stress hormones, which make your heart work harder)
Evaluation Composite: Percent of patients who meet all 4 standards	Treatment Composite: Percent of patients with moderate or severe heart failure who received at least one of the medications

Heart Failure

