



# Using Geospatial Statistics to Analyze Medicaid Data



**Medicaid Innovation  
Accelerator Program  
- Data Analytics  
National Webinar**

***August 31, 2017  
3:00 PM – 4:00 PM EDT***

# Logistics for the Webinar

---

- All lines will be muted
- To participate in a polling question, exit “full screen” mode
- Use the chat box on your screen to ask a question or leave a comment
  - Note: chat box will not be seen in “full screen” mode

# Welcome!

---

- Jessie Parker, GTL and Analyst on Medicaid IAP Data Analytics Team, Data and Systems Group, CMCS

# Today's Speakers

---

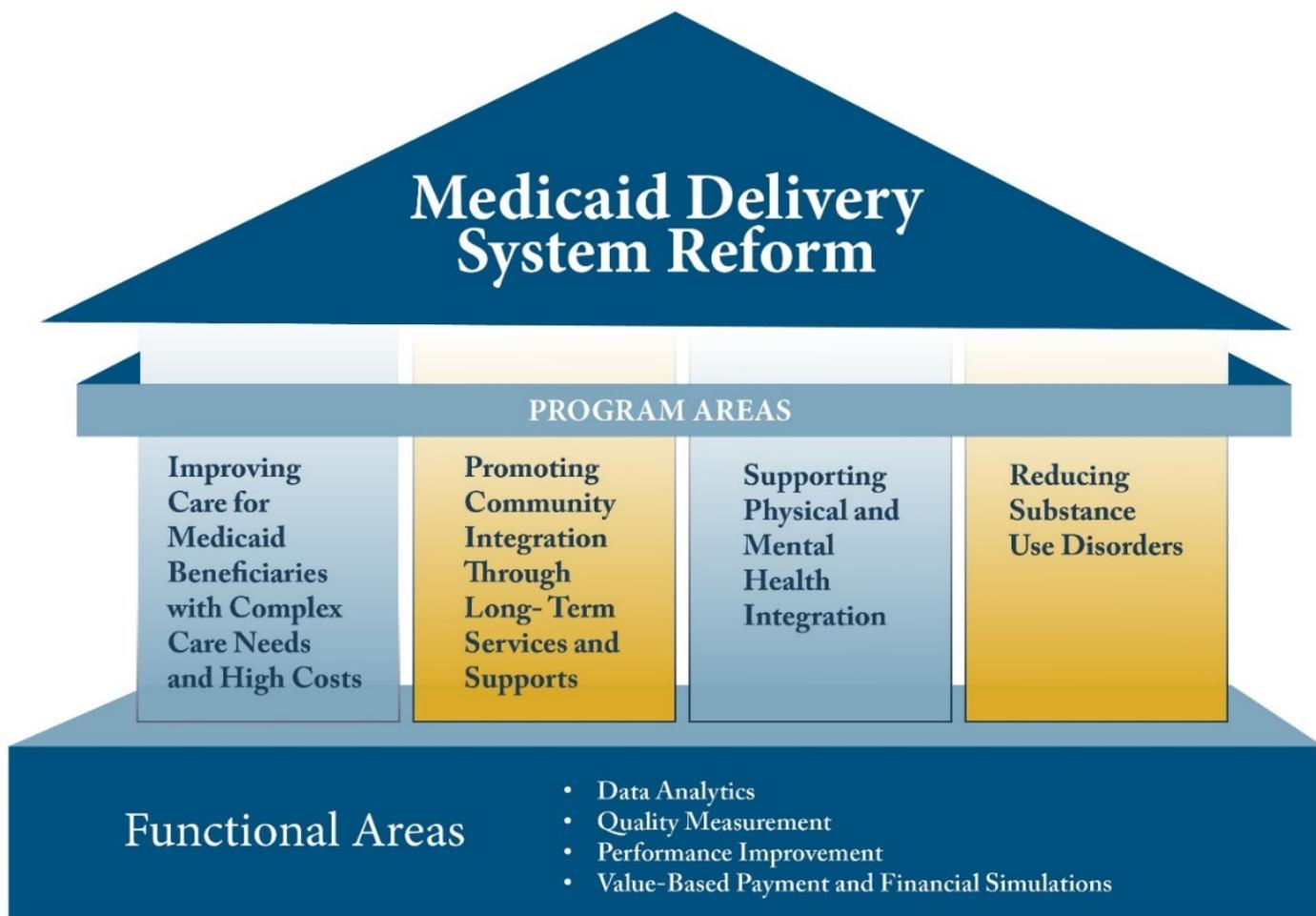
- Marty Jolly, Team Lead Programmer Analyst, Government Health and Human Services, Truven Health Analytics
- Aaron Truchil, Director of Analytics & Informatics, Camden Coalition

# Agenda for Today's Webinar

---

- Overview of Medicaid Innovation Accelerator Program
- Introduction to Approaches for Geospatial Analysis
- Example Geospatial Project
- Overview of Hotspotting Techniques
- Questions and Answers

# Medicaid Innovation Accelerator Program (IAP)



# Goals for Today's Webinar

---

In this interactive webinar, states will learn about:

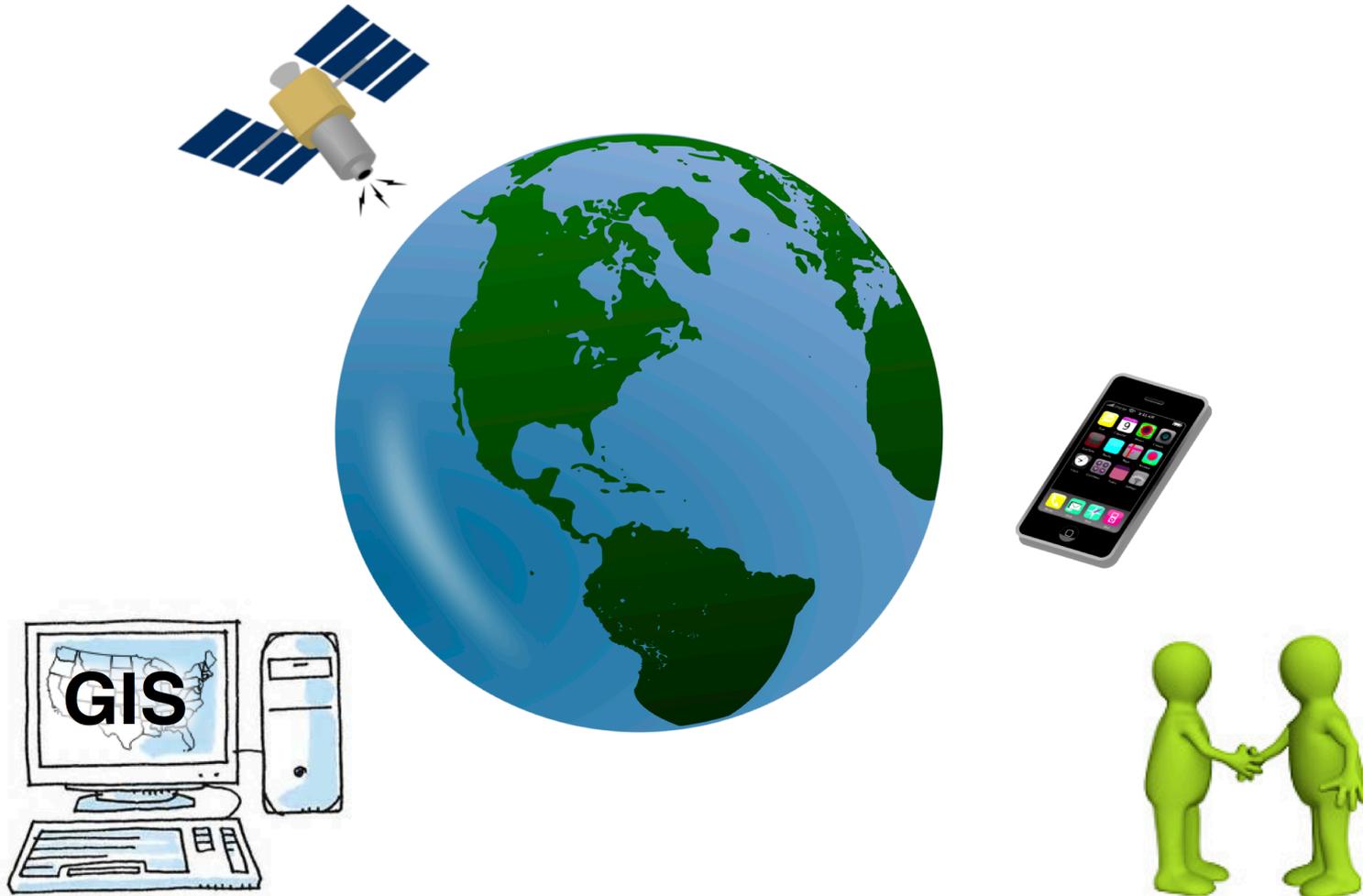
- Map types and usage
- Key questions to ask when evaluating a map
- Geostatistics
- Hotspot maps
- Key components of effective geospatial projects

# Geographic Information Systems (GIS) Foundation

---

**Marty Jolly, Team Lead Programmer Analyst, Government Health and Human Services, Truven Health Analytics**

# Geospatial Analysis - Introduction



# Medicaid Questions

---

## Things on map

- Regions
- Providers
- Population
- Beneficiaries
- Expenditures
- Services

## Questions

- Capacity and Access
  - Network adequacy
  - Accessibility of providers
- Utilization and Expenditure
  - Number of beds
  - Number of provider visits
- Enrollment
- Provider Profiling
- Fraud and Abuse

# Geospatial Team

---

## Roles

- Direction
- Project Management
- Subject Knowledge
- Data Governance
- Computer (GIS) Skills
- Statistical Knowledge

## People

- Stakeholder
- Manager
- PhD statistician
- GIS analyst  
programmer

Emphasis on Data Governance

# Return on Investment

---

Key questions to answer

- Why invest in GIS
- What is level of investment
- When will benefits be delivered
- Who will deliver benefits
- What resources are required
- What is proven financial case

The Business Benefits of GIS: An ROI Approach  
by David Maguire, Victoria Kouyoumijan, Ross Smith  
ESRI Press - 2008

# Example Geospatial Project

---

## The Analysis of Spatial Association by Use of Distance Statistics

by Arthur Getis and J. K. Ord

Geographical Analysis, Vol. 24, Issue 3, pages 199-201

Professor Getis  
San Diego State University



Professor Anselin  
Arizona State University



Professor Ord  
Georgetown University

# North Carolina SIDS

---

- Sudden Infant Death Syndrome, SIDS
- Where: North Carolina
- When: 1979-1984
- Geographic Level: County
- Question: Spatial association of county, SIDS

# NC SIDS - Data

	A	B	C	D	E	F	G
1	NAME	BIR79	SID79	SIDR79			
2	Alamance	5,767	11	1.907404			
3	Alexander	1,683	2	1.188354	NAME	County name	
4	Alleghany	542	3	5.535055	BIR79	County births 1979	
99	Wilson	4,706	13	2.762431	SID79	SIDS deaths 1979	
100	Yadkin	1,568	1	0.637755	SIDR79	(SID79/BIR79) * 1000	
101	Yancey	869	1	1.150748			

## Statistical Data Requirements

- Minimum 30 counties
- Variance among county SIDS counts
- No known spatial association in data
  - e.g. no regional viral outbreaks

NC SIDS 1979

The University of Chicago Center for Spatial Data Science – Sample Data  
<https://s3.amazonaws.com/geoda/data/sids2.zip>

# Data Sources



## TIGER Products

**TIGER** = Topologically Integrated **G**eographic **E**ncoding and **R**eferencing

- [TIGER/Line with Selected Demographic and Economic Data](#)
- Census Bureau data and TIGER products
- State data and GIS portals
- County data and GIS portals

# NC SIDS - Reference Map

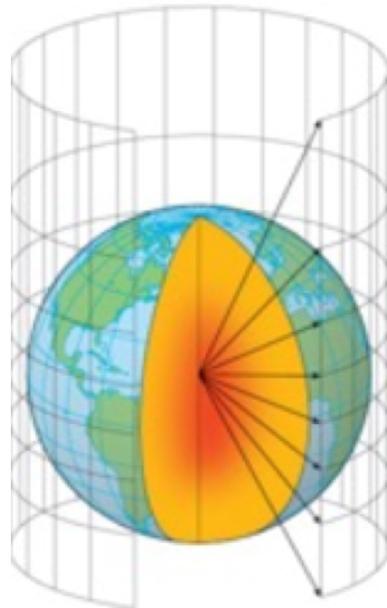


<https://www.census.gov/geo/maps-data/>

# Map Projections

Form post card to ball -  
distortion significant  
for card not for stamp

- Shape
- Area
- Direction
- Distance



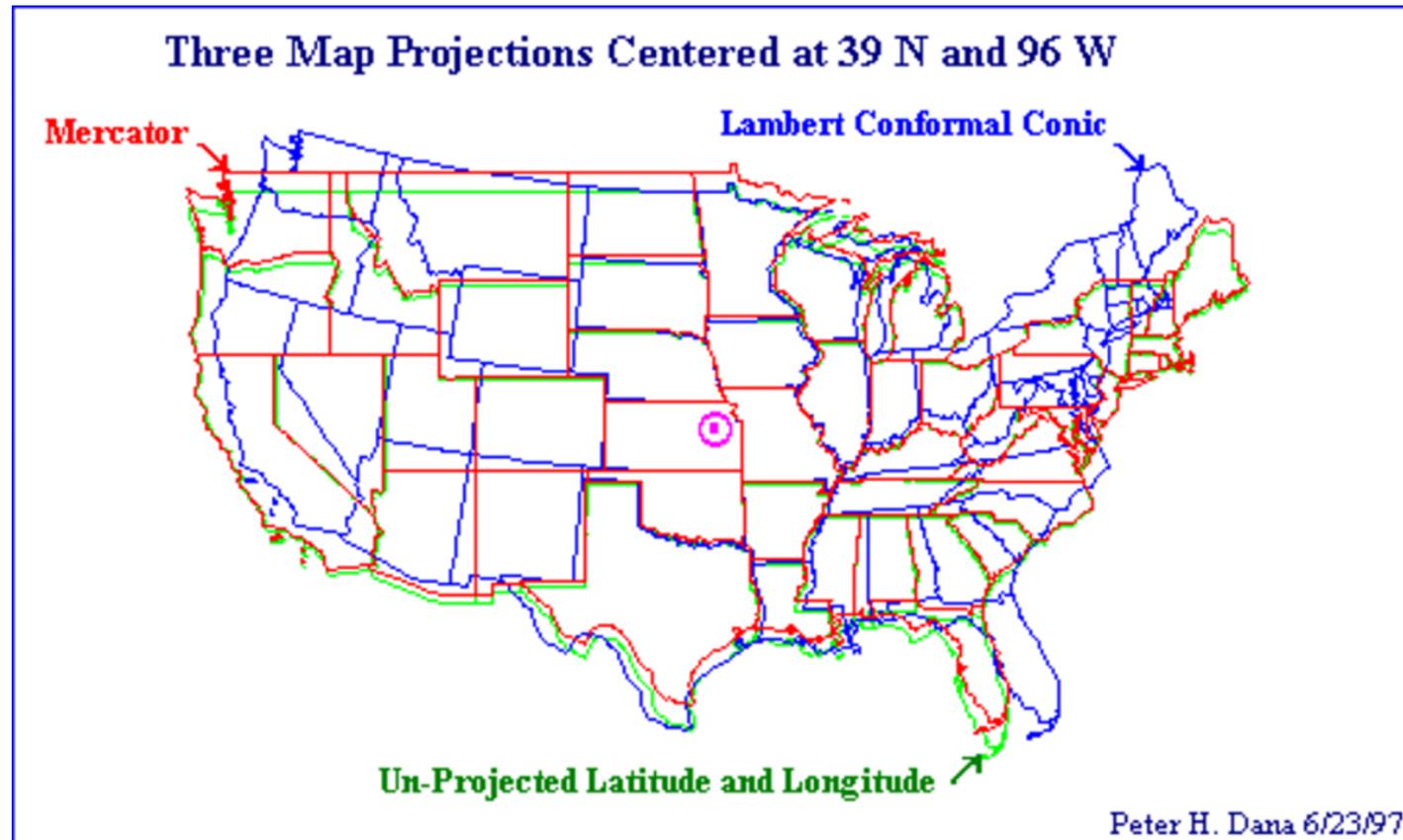
post card ~ national level

stamp ~ state level



<http://www.coxclasses.com/earthscience/ch2/figure5.jpg>

# Projection Comparison National and State Levels

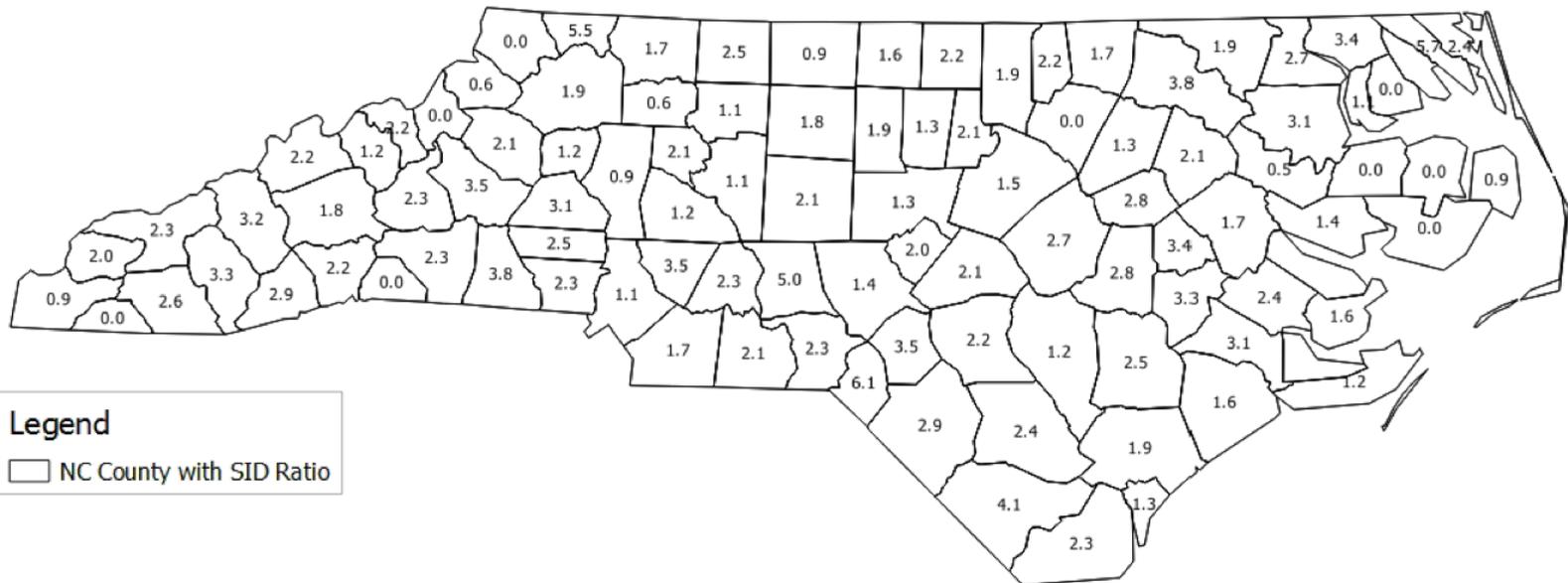


**How to choose a projection, by Jochen Albrecht, Hunter College**

<http://www.geo.hunter.cuny.edu/~jochen/gtech201/lectures/lec6concepts/map%20coordinate%20systems/how%20to%20choose%20a%20projection.htm>

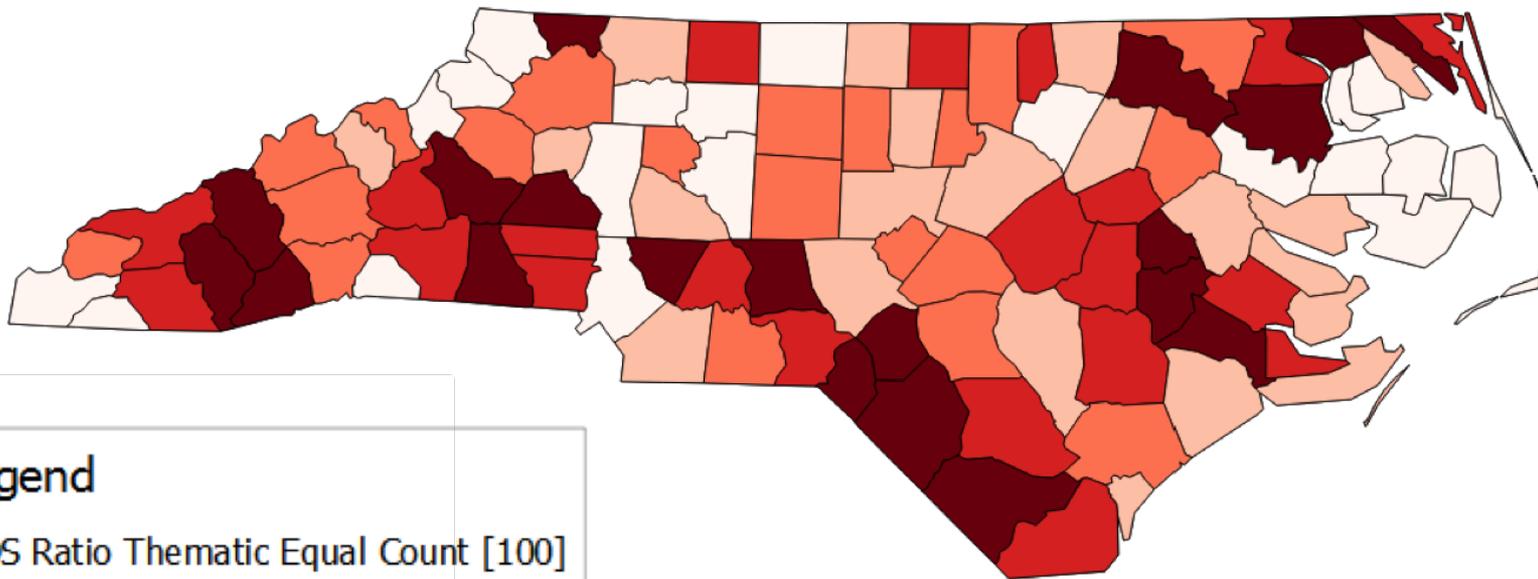
# NC SIDS – Joining Data to Map

NC 1979 SIDS Ratio by County  
SIDS/Births\*1000



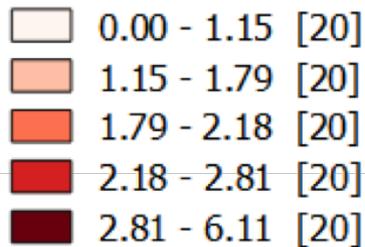
# NC SIDS - Thematic Map Equal Count

NC 1979 SIDS Ratio by County  
SIDS/Births\*1000



## Legend

SIDS Ratio Thematic Equal Count [100]

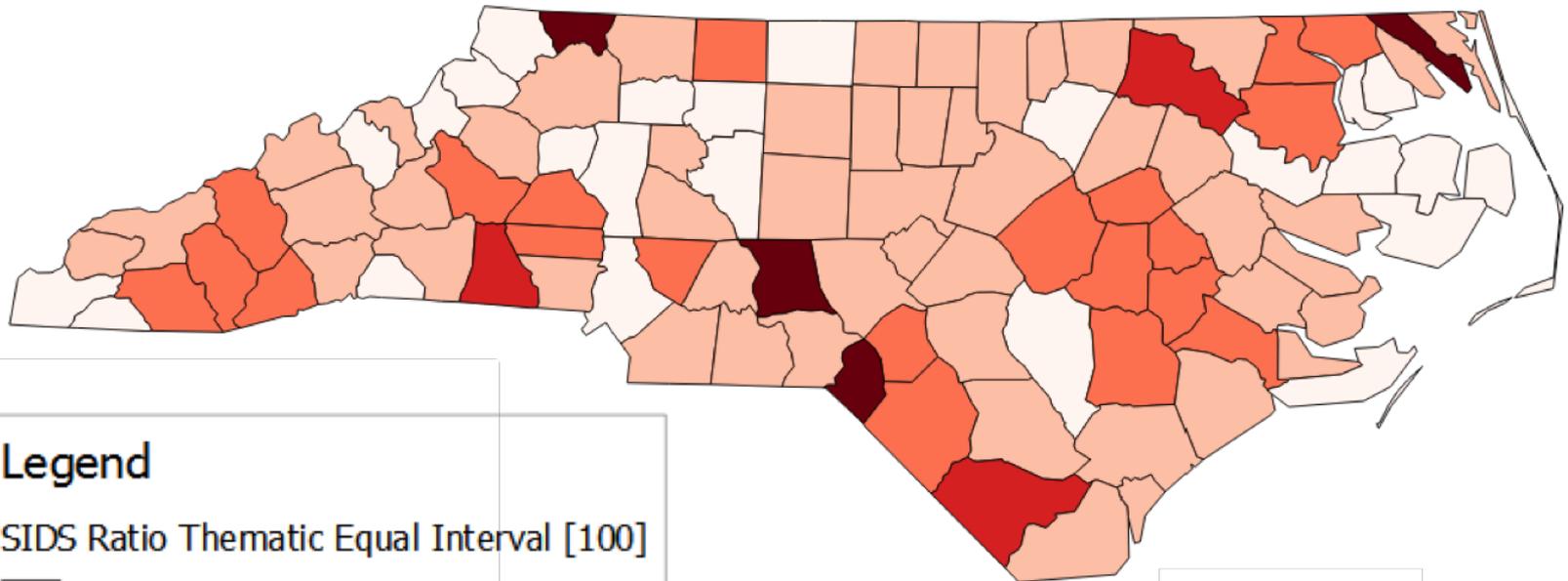


How to Lie with Maps, by Mark Monmonier

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

# NC SIDS - Thematic Map Equal Interval

NC 1979 SIDS Ratio by County  
SIDS/Births\*1000

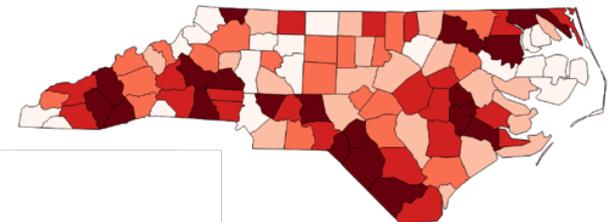


## Legend

SIDS Ratio Thematic Equal Interval [100]

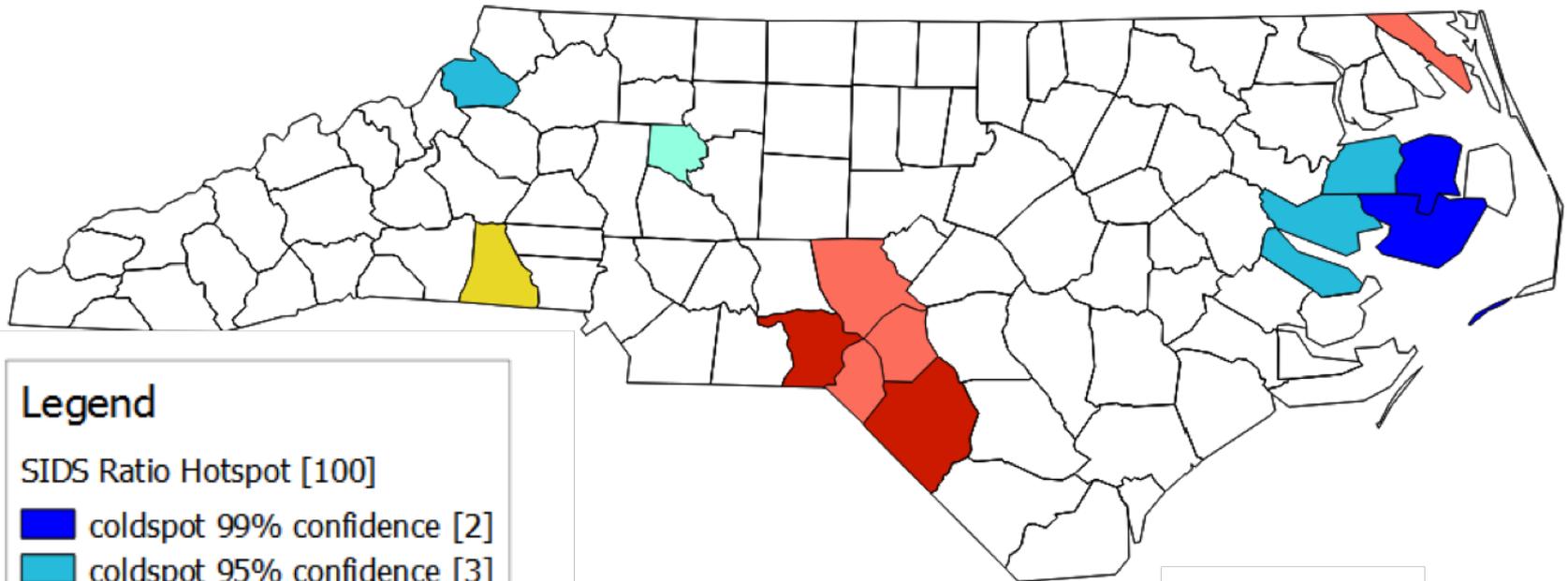
	0.0000 - 1.2228	[24]
	1.2228 - 2.4455	[48]
	2.4455 - 3.6683	[21]
	3.6683 - 4.8911	[3]
	4.8911 - 6.1139	[4]

Equal Count

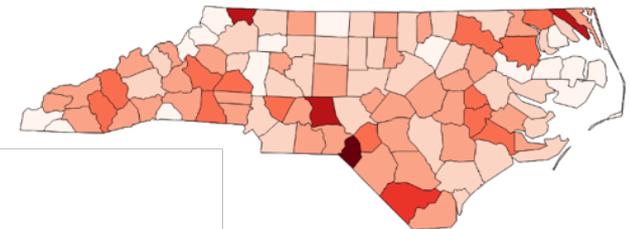


# NC SIDS - Hotspot Map

NC 1979 SIDS Ratio by County  
SIDS/Births\*1000



Getis – Ord Thematic



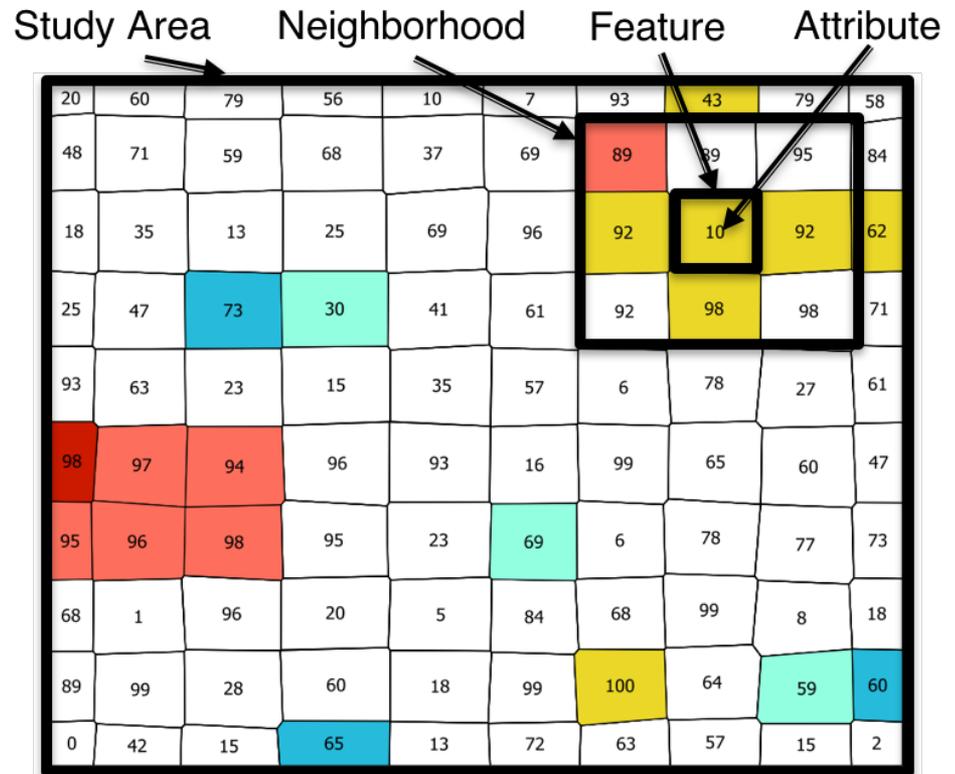
# Hotspot Analysis: Under the Hood

To be a statistically significant hot/cold spot

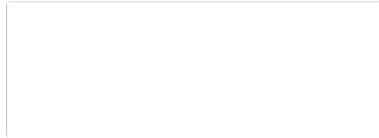
- neighborhood has high/low mean value compared to study area
- high confidence it is not random

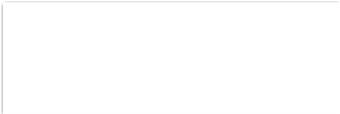
Geospatial Axiom

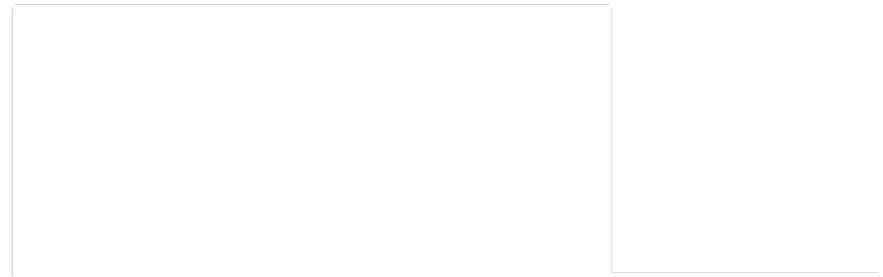
"Everything is related to everything else, but near things are more related than distant things."



# Toolkit



- QGIS
- Excel 
- Internet Resources



# Geospatial Takeaways

---

- Geospatial analysis is iterative
- Options used to create a thematic map can affect conclusions drawn
- Thematic maps are useful but subjective
- Hotspot maps provide statistical rigor, objective analysis
- Data governance is especially important in geospatial projects

# Questions?

---



Camden  
Coalition

Improving Health  
Care & Reducing  
Costs with  
Innovative, Local Data Systems &  
Geospatial  
Analysis

**Aaron Truchil**

Director of Analytics & Informatics | August 2017

# §1 About the Camden Coalition



## MISSION

Spark a field and movement that unites communities of caregivers in Camden and across the nation to improve the wellbeing of individuals with complex health and social needs.



## VISION

A transformed healthcare system that ensures every individual receives whole-person care rooted in authentic healing relationships.



CAMDEN

# Overview of the Camden Coalition

We are a citywide coalition of hospitals, primary care providers, social service providers, and community representatives that collaborate to deliver better healthcare to our most fragile citizens. We innovate and test health care delivery models to improve patient outcomes and reduce the cost of their care using data driven, human-centered approaches.



**Building a Citywide, All-Payer, Hospital Claims Database  
to Improve Health Care Delivery  
in a Low-Income, Urban Community**

Kennen Gross, PhD, MPH,<sup>1</sup> Jeffrey C. Brenner, MD,<sup>1</sup> Aaron Truchil, MS,<sup>1</sup>  
Ernest M. Post, MD,<sup>2</sup> and Amy Henderson Riley, MA, CHES<sup>1</sup>

**Population Health  
Management**

# Camden Hospital Utilization

2014 Snapshot

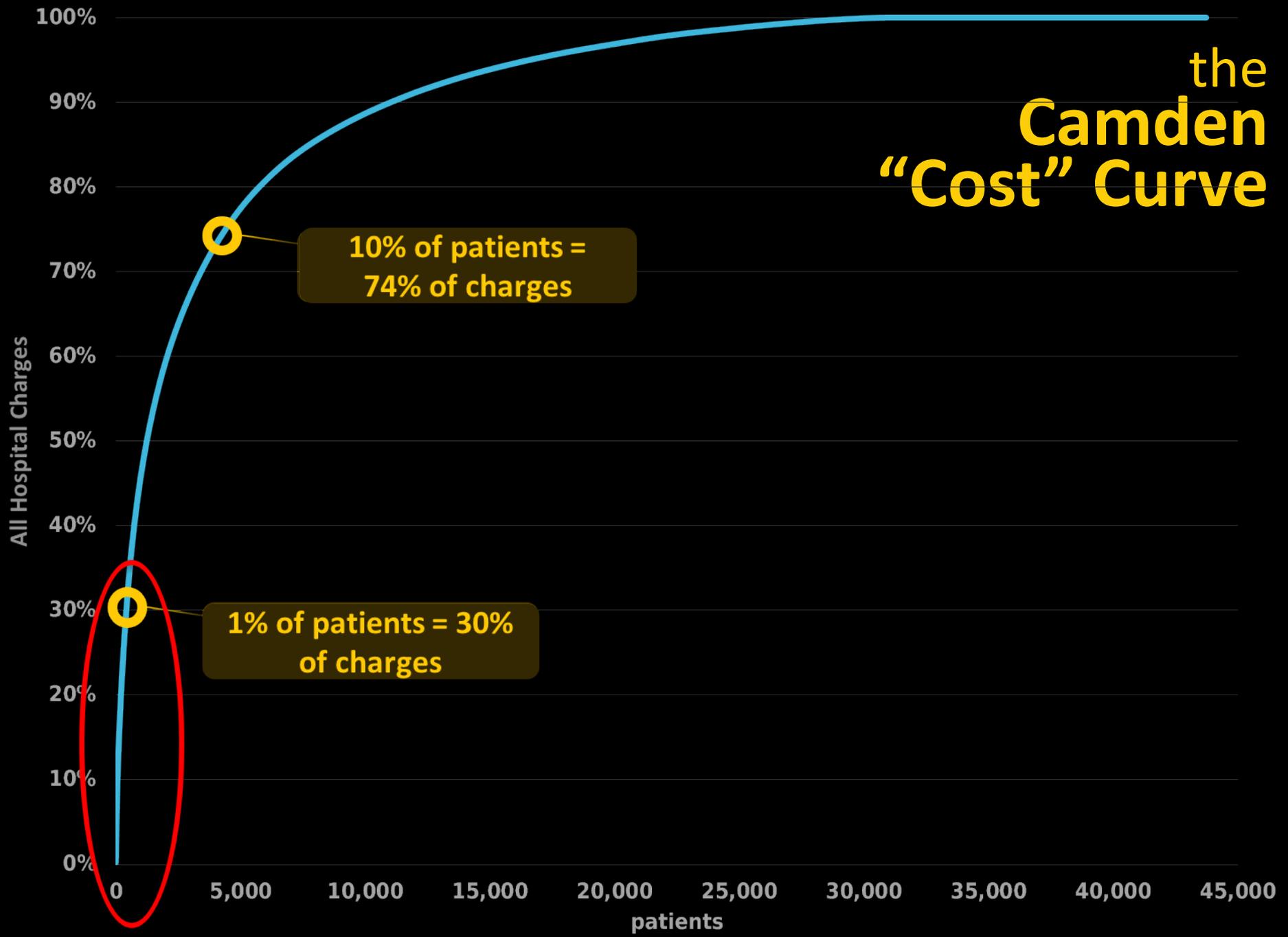
Total Hospital Revenue: **\$132,000,000**

Total Patients with a Hospital Visit: **42,708**

Patients Visiting 2+ Hospitals (Same Year): **23%**

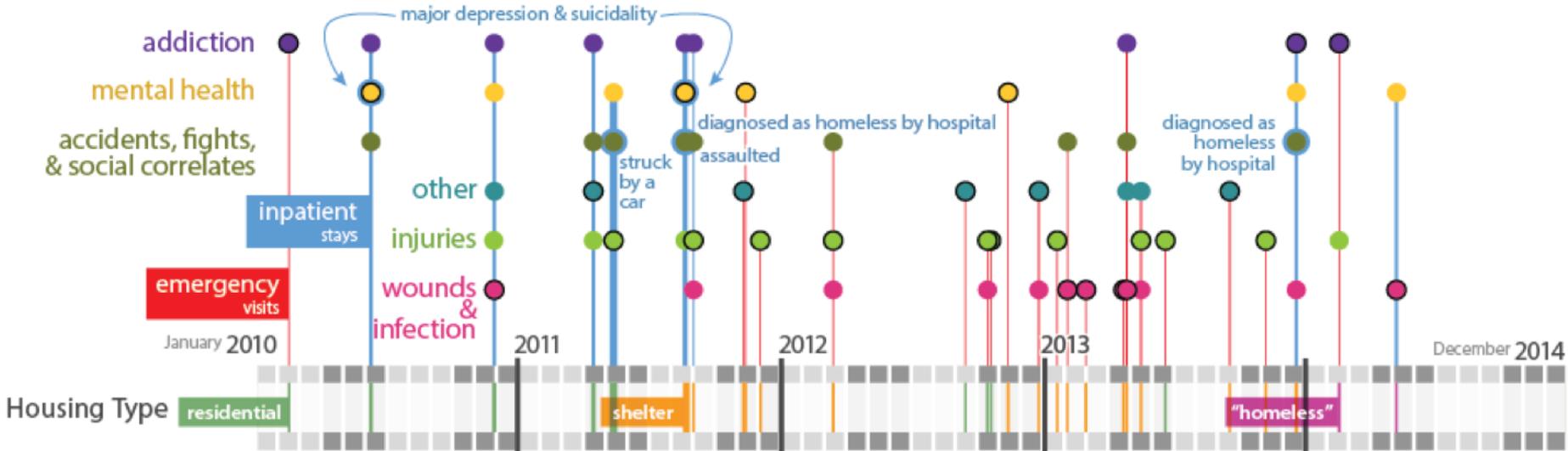
Patients Visiting 2+ Hospitals (Over 5 years): **41%**

# the Camden "Cost" Curve



# Anonymized 1% Case Study

Diagnosis Categories ○ primary diagnosis



# What do Camden's Most Expensive Residents Look Like?

≈1% of population  
>5 chronic conditions

## Averages:

57 years old

4.5 ED visits

5.3 inpatient

Hospitalized 54 days

\$673,000 charges

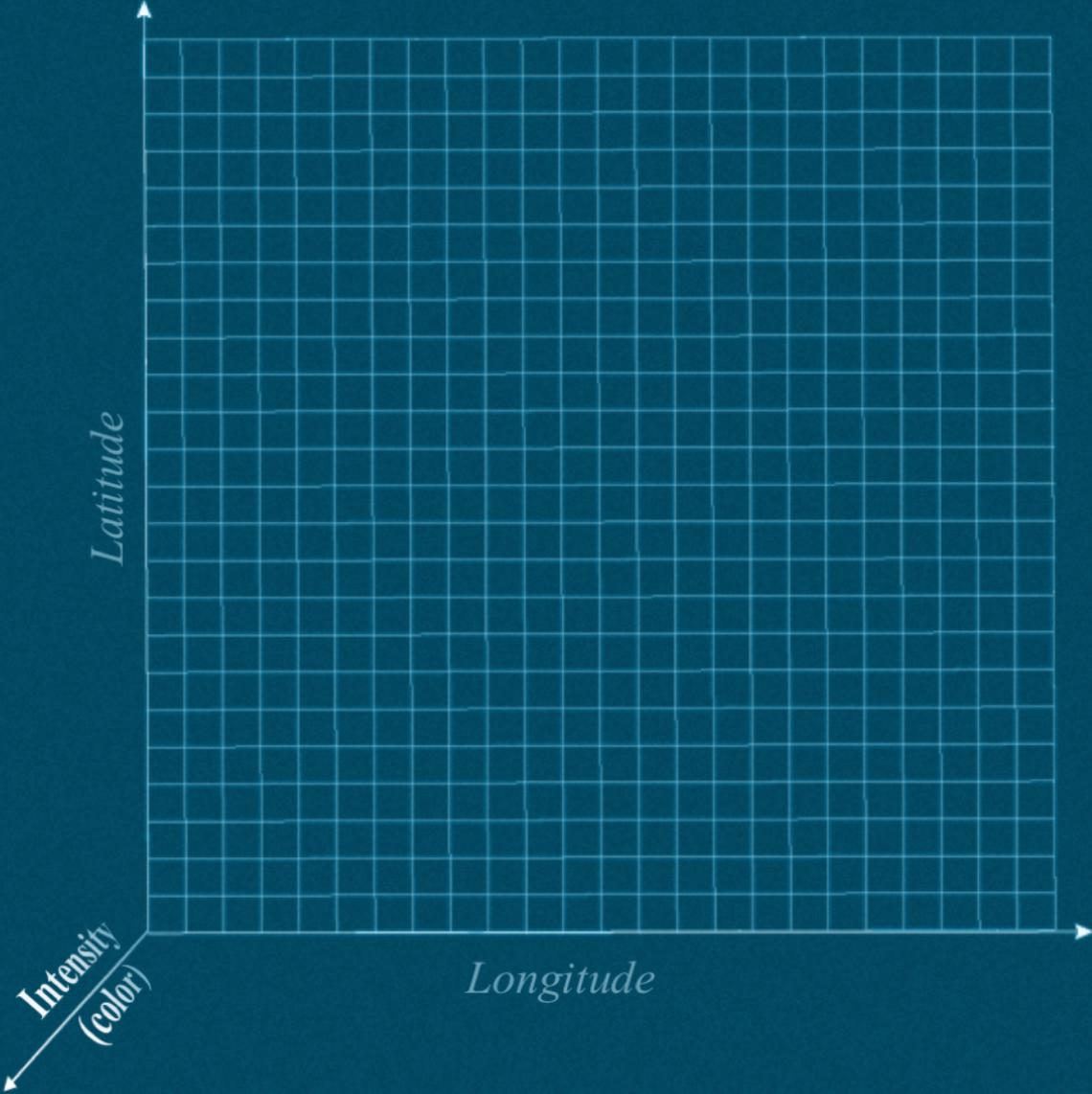
\$73,143 receipts

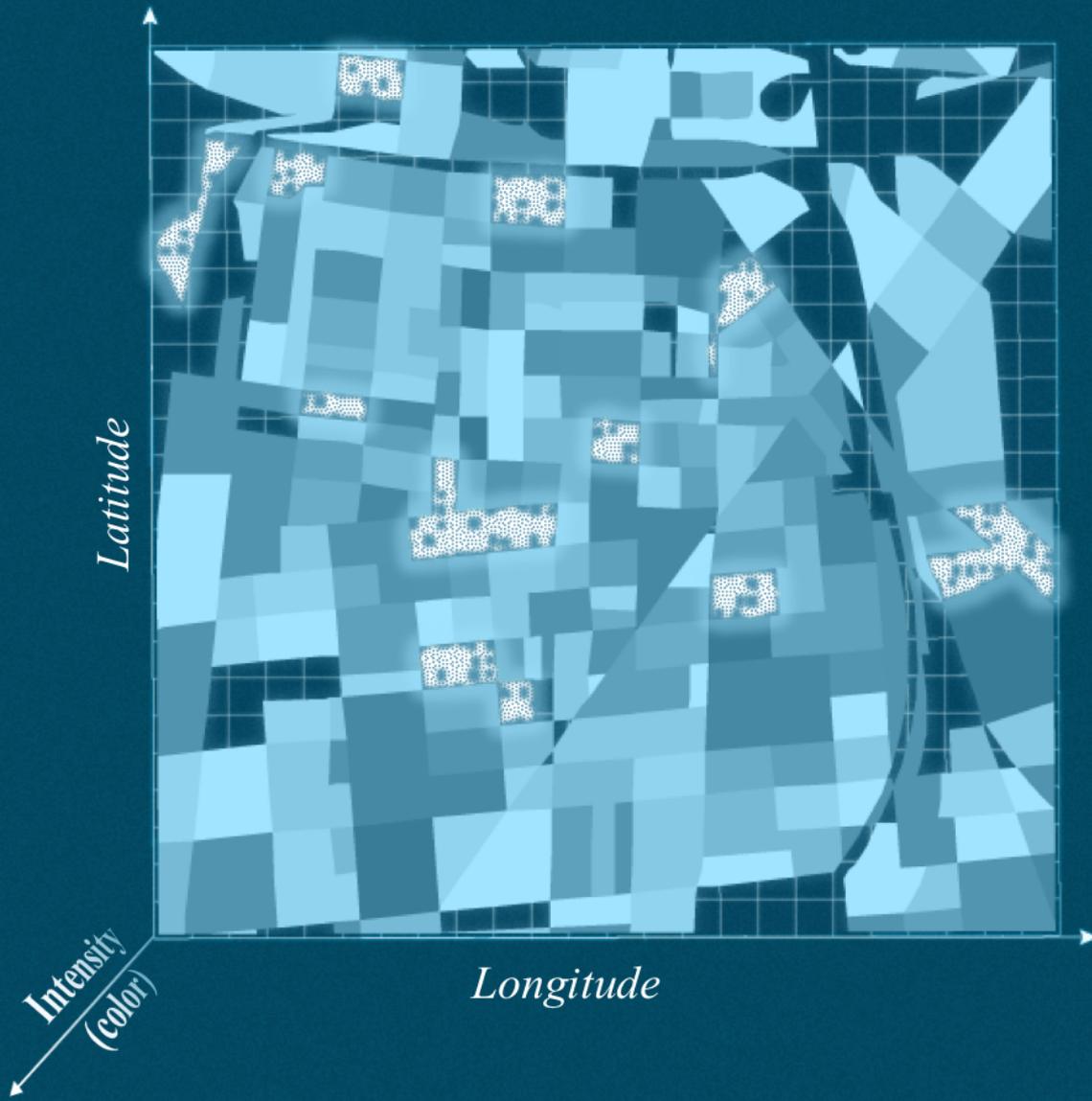
**Healthcare hotspotting** is the strategic use of data to target evidence-based services to complex patients with high utilization.

These patients are experiencing a mismatch between their needs and the services available.

## §2 Incorporating Geospatial Analysis into Hotspotting

ID	Evl	Name	DOB	AdmitDate	Dx_1
414	E	CAIN ANGIE	21-Jan-74	25-Jan-13	574.20
414	E	CAIN ANGIE	21-Jan-74	21-Apr-13	789.09
723	I	LAWRENCE DEBBIE	02-Jun-98	11-Apr-13	614.9
1520	E	CARR SANDY	21-Mar-77	06-Jan-13	942.34
1520	E	CARR SANDY	21-Mar-77	04-Jan-13	788.1
2156	E	GRIFFITH LEROY	06-May-84	12-Apr-13	414.01
2156	E	GRIFFITH LEROY	06-May-84	14-Apr-13	305.00
2156	E	GRIFFITH LEROY	06-May-84	17-Jan-13	414.01
2156	E	GRIFFITH LEROY	06-May-84	30-Jan-13	786.50
2156	E	GRIFFITH LEROY	06-May-84	04-Mar-13	786.59
2156	E	GRIFFITH LEROY	06-May-84	21-Apr-13	786.05
2156	E	GRIFFITH LEROY	06-May-84	28-Jan-13	786.50
2156	E	GRIFFITH LEROY	06-May-84	02-Apr-13	305.00
2283	I	VASQUEZ EDDIE	02-Apr-65	06-Apr-13	923.3
2283	E	VASQUEZ EDDIE	02-Apr-65	15-Apr-13	996.01
2283	E	VASQUEZ EDDIE	02-Apr-65	26-Jan-13	996.01
2696	I	STONE JESSE	10-Aug-07	13-Jan-13	913.0
2696	I	STONE JESSE	10-Aug-07	25-Mar-13	943.02
2893	E	ALLEN KENNETH	06-Jul-95	04-Mar-13	305.90
3087	E	NEWMAN BERNICE	06-Sep-07	09-Feb-13	643.03
3555	I	VEGA JULIO	08-Jul-51	29-Mar-13	842.00
4628	E	CHAMPMAN LINDSEY	29-Aug-66	06-Jan-13	845.09
4808	E	WALTON MELANIE	18-Jul-92	21-Jan-13	499





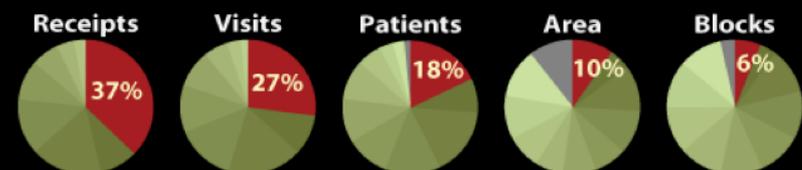
# Where do Camden's Most Expensive Residents Reside?

Several buildings (e.g.) annually generate \$1-\$3 million in hospital costs.

6% of city blocks account for 18% of patients and 37% of receipts.



January 2002  
– June 2008





## Example High Cost Building

### Northgate II

Over 5 years...

615 patients accounted for 3,901 hospital visits \$12 million in hospital receipts

### Top Diagnoses when Visiting the Hospital

#### Emergency Room

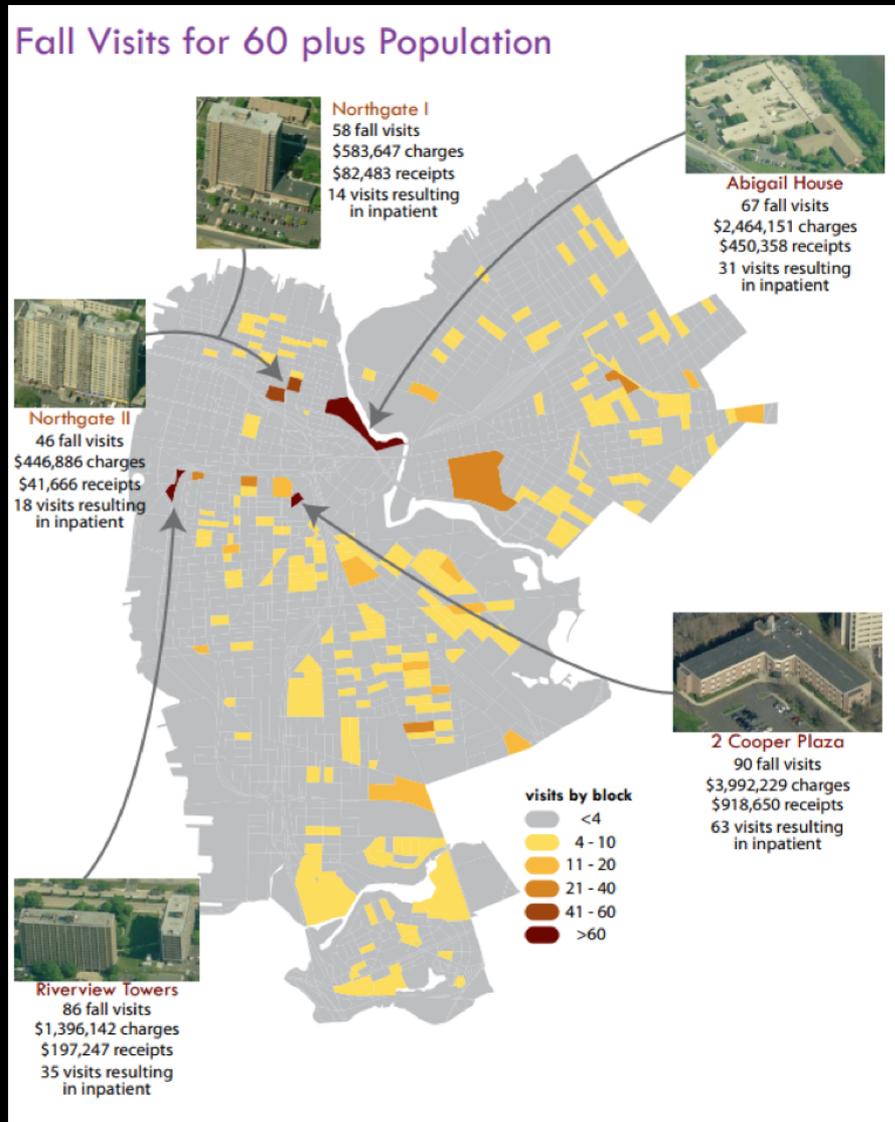
1. Diabetes
2. Lung and chest symptoms (colds, flu)
3. General symptoms
4. Hypertension
5. Stomach and pelvis symptoms

#### Inpatient

1. Hypertension
2. Diabetes
3. Heart Disease
4. Heart Failure
5. Asthma



# Other Uses of Geospatial Analyses



# Other Uses of Geospatial Analyses (cont.)

## Pregnancies by Neighborhood in Camden

Figure 11 Average Yearly Deliveries by Neighborhood

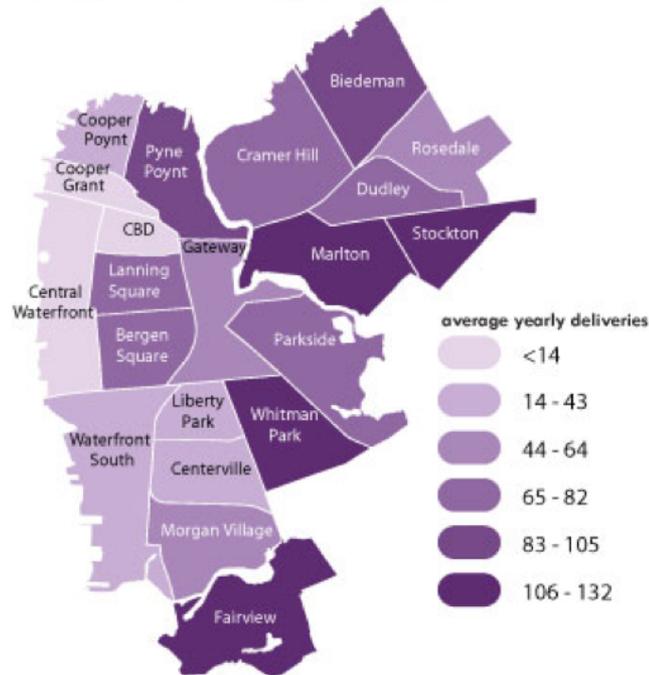
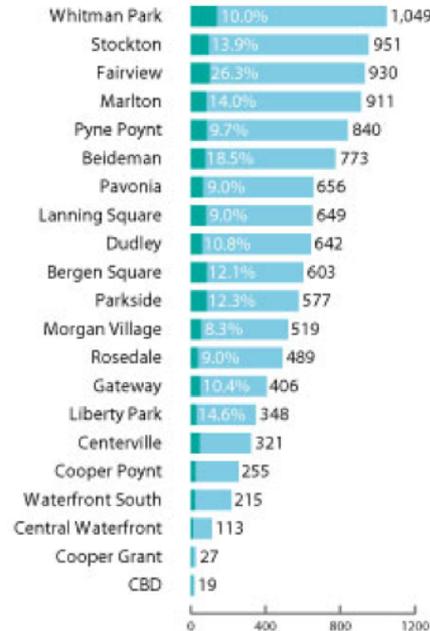


Figure 12 Total Deliveries (Preterm and Normal) by Neighborhood (2002 - 2009)



## The Cost of Preterm Pregnancies

Table 4 Total Cost of Pregnancy for Normal Delivery and Preterm Delivery\*

	Normal Delivery			Preterm Delivery			Difference (Preterm - Normal)		
	LOS	Charges**	Receipts	LOS	Charges	Receipts	LOS	Charges	Receipts
avg	2.3	\$24,521	\$4,434	19.7	\$173,682	\$18,318	17.4	\$149,161	\$13,884
median	2	\$20,359	\$4,060	11	\$82,473	\$7,776	9	\$62,114	\$3,716

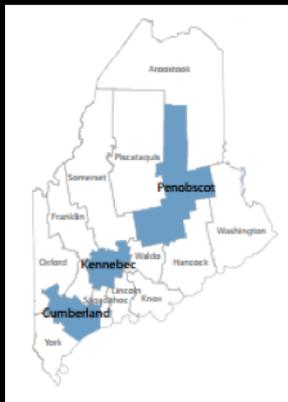
\*Normal delivery and preterm delivery were determined by baby's initial length of stay. Babies with an initial length of stay greater than 4 days were categorized as preterm deliveries. Using this formula, we came to 12.7% of all deliveries, very close to the 12.1% estimated by NJ Vital Statistics.

\*\* Charges reflect the amount charged by the Hospital to the Payer for a particular visit. Receipts reflect the amount that was actually collected.

# §3 Hotspotting Outside of Camden

# MaineCare Hotspotting Analysis

An analysis of 2 years of  
Medicaid claims data for 3  
Maine counties (Cumberland,  
Kennebec, and Penobscot)



Report available at:

[https://www1.maine.gov/dhhs/oms/pdfs\\_doc/vbp/CCHP\\_04062012\\_MaineCare\\_Report.pdf](https://www1.maine.gov/dhhs/oms/pdfs_doc/vbp/CCHP_04062012_MaineCare_Report.pdf)

## MaineCare Hospital Utilization Analysis for Cumberland, Kennebec, and Penobscot Counties, FY2009 & FY2010

### About the Project

The Camden Coalition of Healthcare Providers (CCHP) is a nine-year old strategic initiative with a mission to improve the quality, capacity, and accessibility of the healthcare system for vulnerable populations in the City of Camden. The Camden Coalition of Healthcare Providers has compiled the Camden Health Database, a citywide all-payer, all-provider claims database that contains data on all hospital encounters from 2002 through 2010. The Camden Health Database has shown to be a tremendous tool for quantifying and analyzing local health trends. Using its expertise in managing and analyzing claims data, CCHP has analyzed 2 years of Medicaid claims data from the MaineCare database. Data was extracted for three counties (Cumberland, Kennebec, and Penobscot) for the 2009 and 2010 fiscal years.

### Summary of Findings

For the study area in 2009, MaineCare paid \$123.7 million for 73,821 ED visits and 12,877 Inpatient (IP) visits made by 38,485 unique patients. For the study area in 2010, MaineCare paid \$136.8 million for 78,723 ED visits and 12,880 Inpatient visits made by 41,339 unique patients.

614 (1%) of patients accounted for 31.6% of total hospital costs during the 2 year period ; 12,228 (20%) patients accounted for 87% of costs during the 2 year period.

High utilizer patients are defined as those patients with 6 or more ED visits and/or 3 or more IP visits during the 2 year time period. 6,121 patients (9.9%) met this "High Utilizer" definition. While High Utilizers represented less than 10% of all MaineCare patients they accounted for 46% of all hospital costs.

### Inpatient High Utilizers

The three most prevalent inpatient diagnosis for High Utilizers were "alcohol-related disorders", "mood disorders", and "chronic obstructive pulmonary disease and bronchiectasis". Inpatient High Utilizers are 2.12 times more likely to have an Inpatient stay with a diagnosis of "alcohol-related disorders" compared to non-High Utilizers, 1.97 times more likely to have an inpatient stay with a diagnosis of "chronic obstructive pulmonary disease and bronchiectasis", and 1.81 times more likely to have a diagnosis of "diabetes" compared to non-High Utilizers. 72% of all IP High Utilizers were over age 34

### ED High Utilizers

The three most prevalent emergency department diagnosis for High Utilizers were "sprains and strains", "disorders of teeth and jaw", and "other upper respiratory infections". ED High Utilizers are 1.46 times more likely to have an ED visit with a diagnosis of "anxiety disorders" compared to non-High Utilizers, 1.46 times more likely to have an ED visit with a diagnosis of "spondylosis; intervertebral disc disorders; other back problems", and 1.38 times more likely to have a diagnosis of "Headache; including migraine". 67% of all ED High Utilizers were under age 35

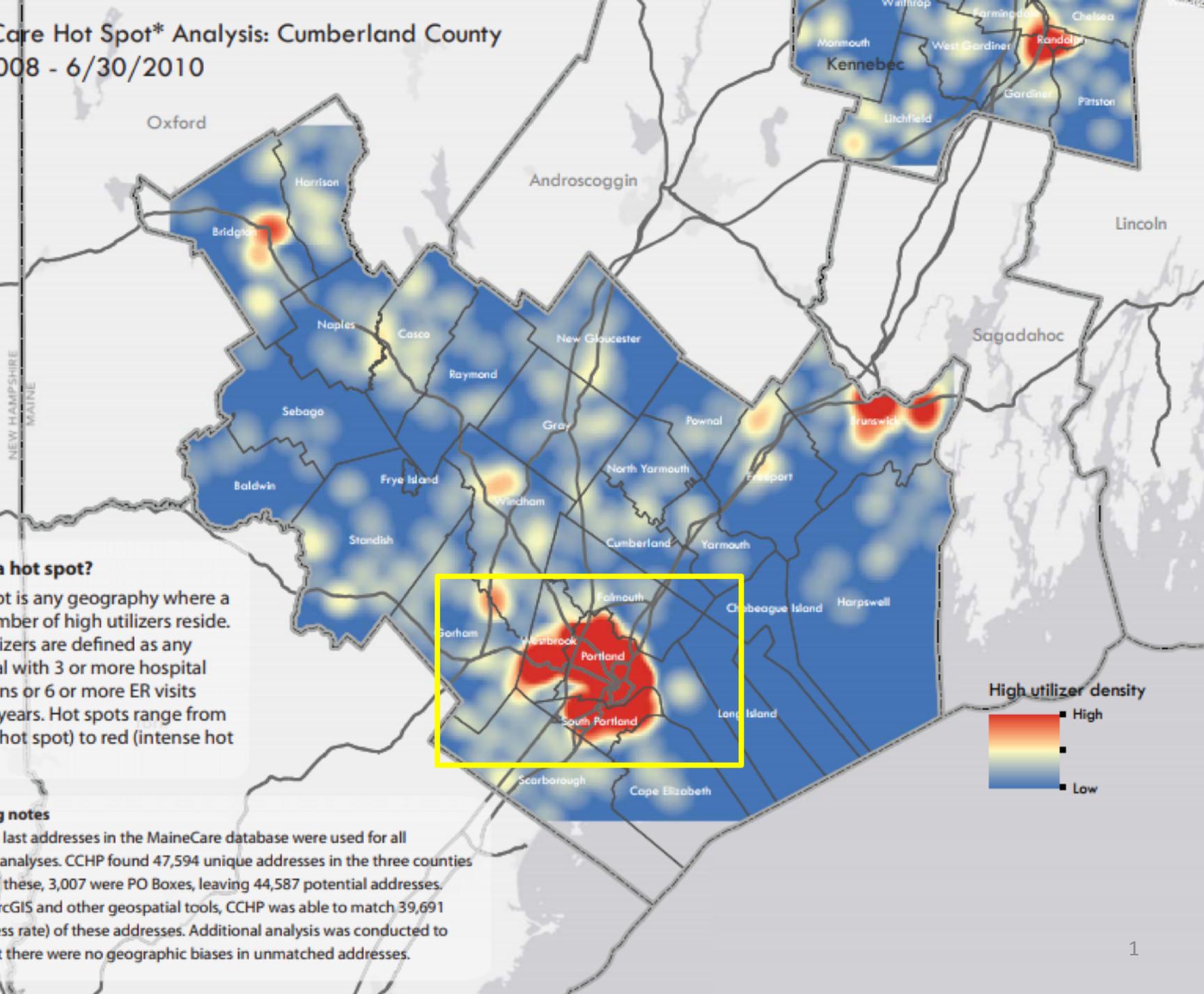
Portland (993), Bangor (462), Waterville (426) and Augusta (357) had the highest prevalence of High Utilizer patients. Together, these four towns contain 46% of all high utilizers. Of all towns with at least 200 MaineCare members, Waterville (14.95%), Lincoln (12.78%), and Winslow (12.25%) had the highest rate of High Utilizers.



Camden Coalition of  
Healthcare Providers  
www.camdencc.org 47

# MaineCare Hot Spot\* Analysis: Cumberland County

7/1/2008 - 6/30/2010



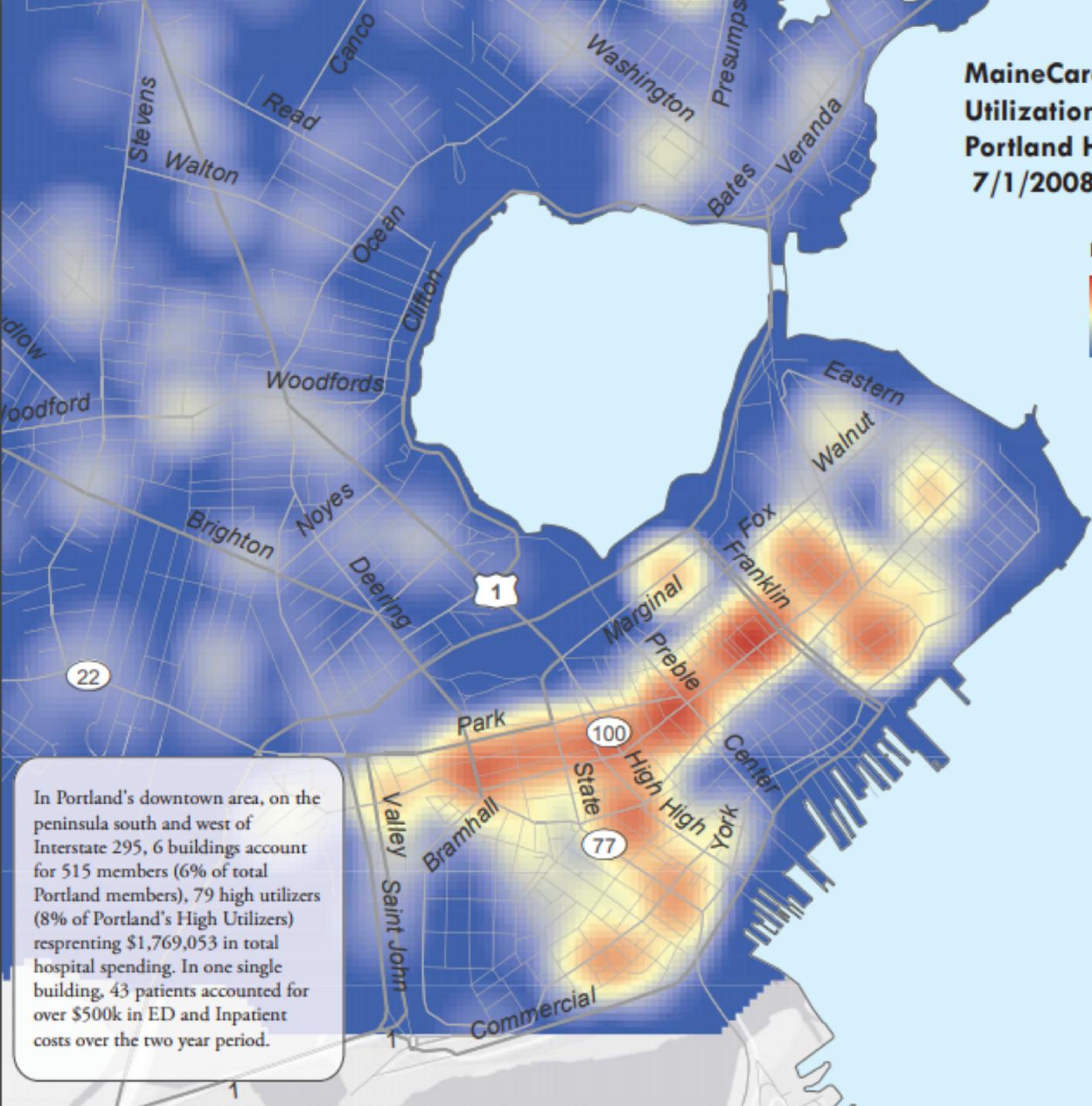
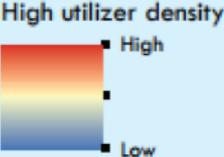
## What is a hot spot?

A hot spot is any geography where a large number of high utilizers reside. High Utilizers are defined as any individual with 3 or more hospital admissions or 6 or more ER visits within 2 years. Hot spots range from blue (no hot spot) to red (intense hot spot)

## Geocoding notes

Individuals last addresses in the MaineCare database were used for all geospatial analyses. CCHP found 47,594 unique addresses in the three counties studied. Of these, 3,007 were PO Boxes, leaving 44,587 potential addresses. Through ArcGIS and other geospatial tools, CCHP was able to match 39,691 (89% success rate) of these addresses. Additional analysis was conducted to ensure that there were no geographic biases in unmatched addresses.

**MaineCare Hospital  
Utilization Analysis  
Portland Hot Spots\*  
7/1/2008 - 6/30/2010**



In Portland's downtown area, on the peninsula south and west of Interstate 295, 6 buildings account for 515 members (6% of total Portland members), 79 high utilizers (8% of Portland's High Utilizers) representing \$1,769,053 in total hospital spending. In one single building, 43 patients accounted for over \$500k in ED and Inpatient costs over the two year period.

Town	Total members that reside in town	# of inp visits from residents of town	# of ED visits from residents of town	Percent of town's members that are High Utilizers	This town has what percentage of all members	This town has what percentage of all High Utilizers	This town has what percentage of all inpatient visits	This town has what percentage of all ER visits
Enfield	129	28	371	15.50%	0.26%	0.41%	0.17%	0.31%
Waterville	2,849	834	8,957	14.95%	5.71%	8.73%	4.94%	7.43%
Veazie	97	37	241	14.43%	0.19%	0.29%	0.22%	0.20%
Lagrange	97	20	243	13.40%	0.19%	0.27%	0.12%	0.20%
Lincoln	947	256	2,648	12.78%	1.90%	2.48%	1.52%	2.20%
Winslow	939	286	2,473	12.25%	1.88%	2.36%	1.70%	2.05%
Newport	471	133	1,310	12.10%	0.94%	1.17%	0.79%	1.09%
Clifton	109	34	235	11.93%	0.22%	0.27%	0.20%	0.20%
Portland	8,360	3,007	23,728	11.88%	16.76%	20.35%	17.82%	19.69%
Pownal	59	18	124	11.86%	0.12%	0.14%	0.11%	0.10%
Brunswick	1,481	674	3,654	11.82%	2.97%	3.59%	3.99%	3.03%
Plymouth	195	39	540	11.79%	0.39%	0.47%	0.23%	0.45%
Alton	85	39	232	11.76%	0.17%	0.20%	0.23%	0.19%
Benton	391	114	973	11.76%	0.78%	0.94%	0.68%	0.81%
Oakland	786	195	2,089	11.70%	1.58%	1.89%	1.16%	1.73%
Millinocket	712	178	1,898	11.66%	1.43%	1.70%	1.05%	1.57%
Bradley	88	26	232	11.36%	0.18%	0.20%	0.15%	0.19%
Medway	108	34	271	11.11%	0.22%	0.25%	0.20%	0.22%
Augusta	3,237	1,158	8,235	11.03%	6.49%	7.32%	6.86%	6.83%
Clinton	399	113	1,052	11.03%	0.80%	0.90%	0.67%	0.87%
Casco	328	100	758	10.98%	0.66%	0.74%	0.59%	0.63%
Bangor	4,490	1,721	11,084	10.29%	9.00%	9.47%	10.20%	9.20%
Vassalboro	391	102	983	10.23%	0.78%	0.82%	0.60%	0.82%
Bridgton	560	182	1,331	10.18%	1.12%	1.17%	1.08%	1.10%
Chester	69	22	183	10.14%	0.14%	0.14%	0.13%	0.15%

# Final Takeaways

- GIS: one of many tools in your analytic toolbox
- Geospatial analysis is not limited to high utilizers
- Start with what's relevant to your existing work
- Building out GIS capacity doesn't have to be complicated and/or expensive
  - Low cost & open source software
  - Ability to train existing analytics staff

# Thank You!



Camden  
Coalition

[aaron@camdenhealth.org](mailto:aaron@camdenhealth.org)

# Final Questions?

---

# Key Takeaways

---

- Benefits of geospatial analysis include:
  - Relevancy to Medicaid questions on capacity, access, utilization, and expenditures
  - Useful for identifying areas of interest for further analysis
  - Relatively low cost when executed by data analysts with open source software

# Survey

---

Thank you for joining today's webinar!

Please take a moment to complete  
the post-webinar survey -

We appreciate your feedback!