Sub-County Statistical Analysis and Visualization using ArcGIS Pro and Python

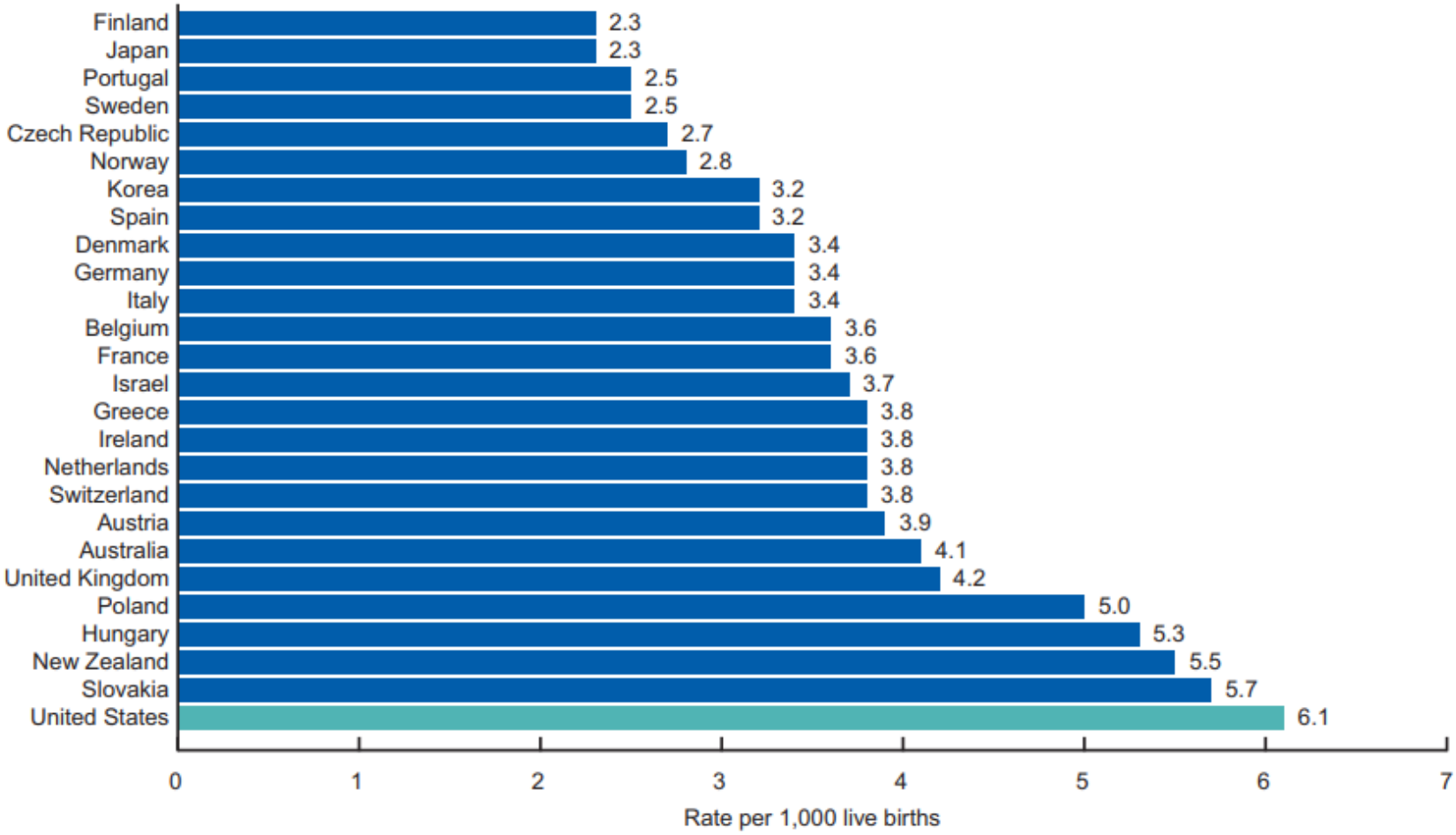
Indiana GIS Day
September 17, 2019

Robert Gottlieb, *GIS Data Analyst*
Epidemiology Resource Center
Jenny Durica, *Director of MCH Epidemiology*
Division of Maternal and Child Health

Indiana State Department of Health

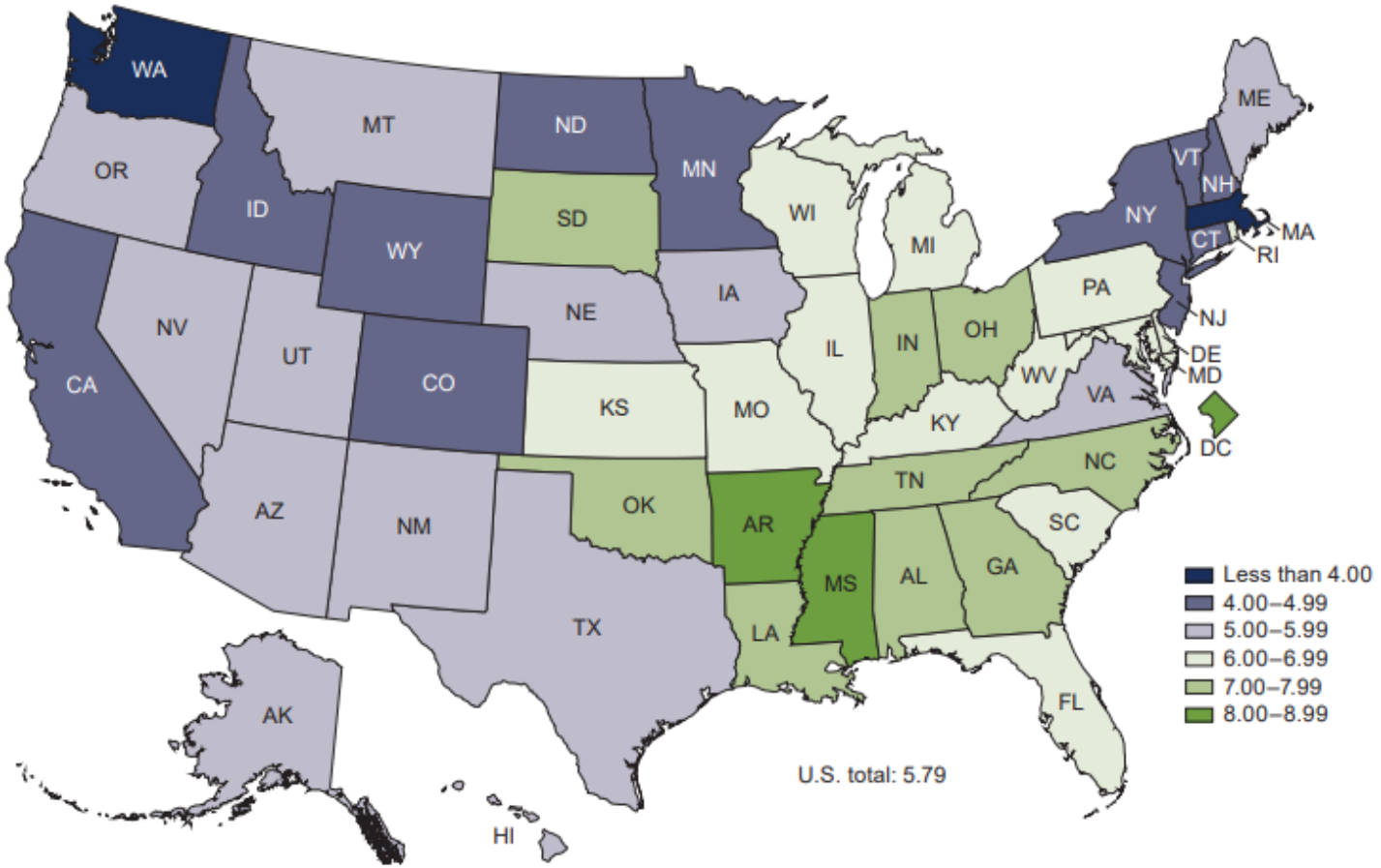
Infant Mortality Rates

U.S. & Selected Countries, 2010



Infant Mortality Rates

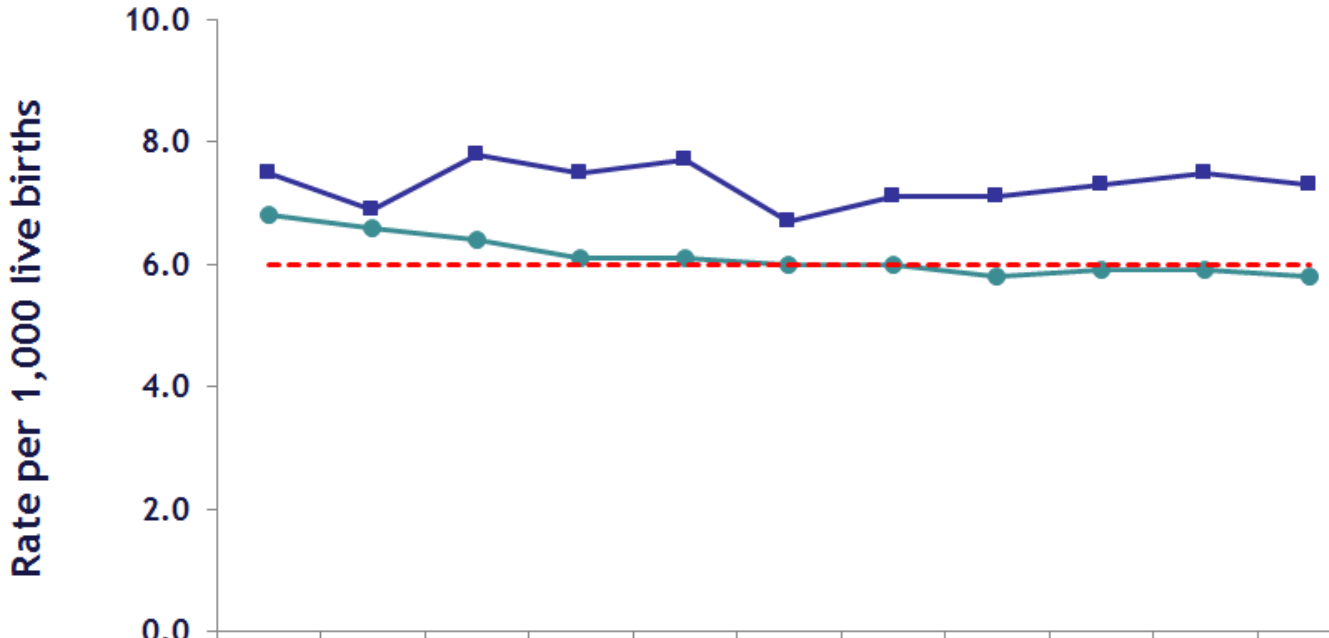
U.S., 2017



SOURCE: NCHS, National Vital Statistics System, Linked birth/infant death file.

Infant Mortality Rates

Indiana, U.S. & Healthy People 2020 Goal, 2007 – 2017



	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Indiana	7.5	6.9	7.8	7.5	7.7	6.7	7.1	7.1	7.3	7.5	7.3
U.S.	6.8	6.6	6.4	6.1	6.1	6.0	6.0	5.8	5.9	5.9	5.8
HP 2020 Goal	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0

ISDH Programs and Initiatives to Help Reduce Infant Mortality Rates

Indiana's Early Start Program

DOSE
Direct On Scene Education

INDIANA PRAMS
Pregnancy Risk Assessment Monitoring System

Baby & Me
Tobacco Free
Healthy Babies. Born on Time.

NEWBORN SCREENING INDIANA

IPQIC
Indiana Quality Improvement Collaborative

INDIANA child fatality review
prevention through understanding...

MCH MOMS HELPLINE
1-844-MCH-MOMS
(844-624-6667)
The key to a healthy baby and a happy mom

Liv

Safe Sleep Program

INDIANA FIMR

Labor of Love
Helping Indiana Reduce Infant Death

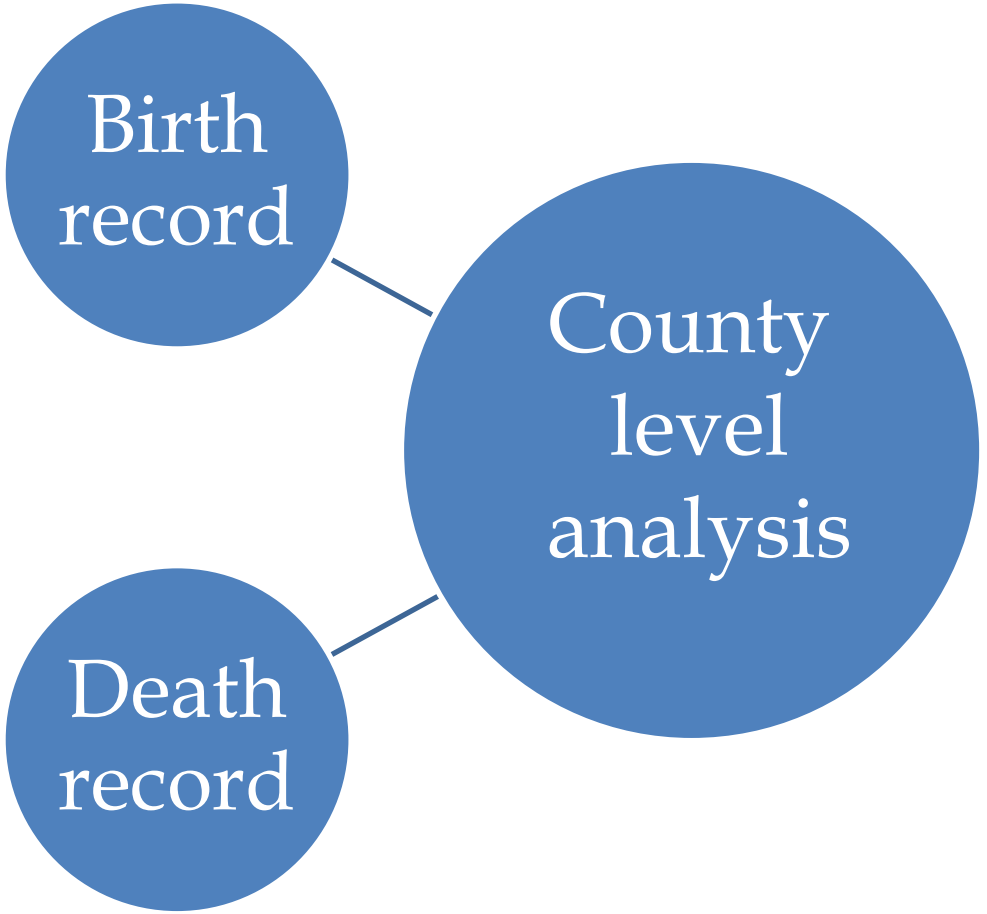
BDPR

Nurse-Family Partnership
Helping First-Time Parents Succeed

Help Me Grow Indiana

Infant Mortality Data

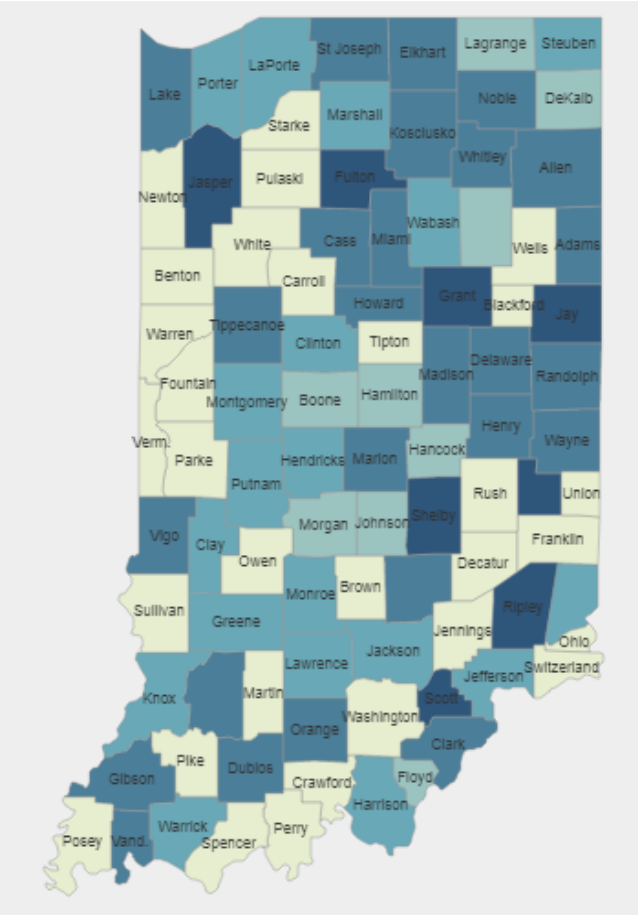
- Pre-term, low birthweight
- Prenatal care
- Smoking during pregnancy
- Insurance



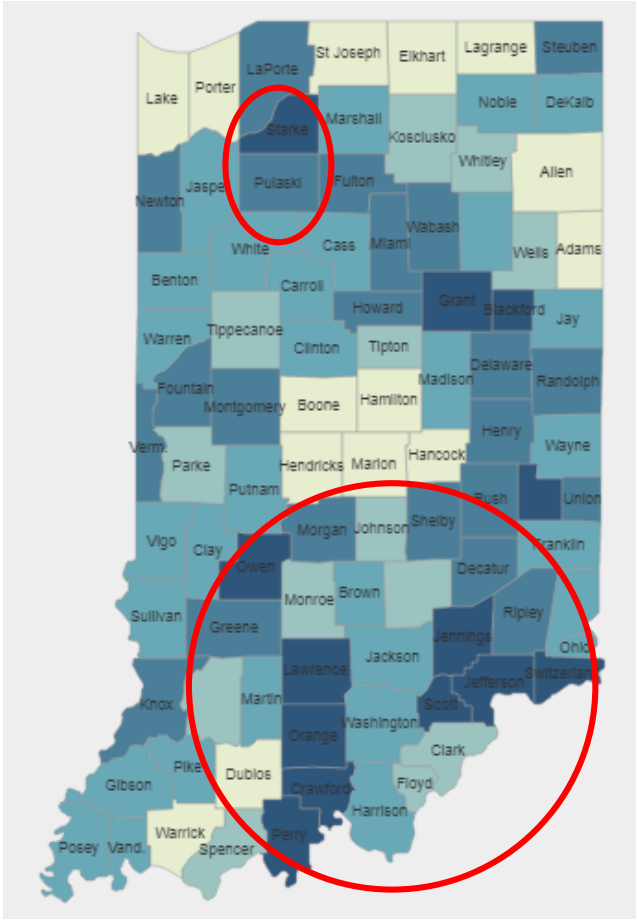
- Cause of death
- Age at death

Infant Mortality and Birth Risk Factors

Infant Mortality Rates, 2013-17



Mothers Smoking During Pregnancy, 2013-17

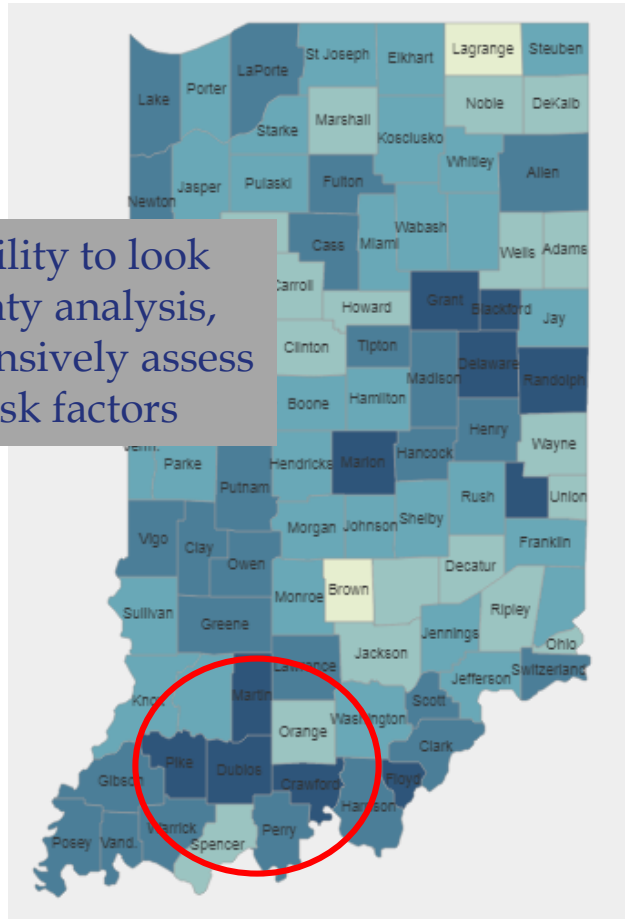


Infant Mortality and Birth Risk Factors

Infant Mortality Rates, 2013-17

Need the ability to look beyond county analysis, and comprehensively assess multiple risk factors

Preterm Infants, 2013-17

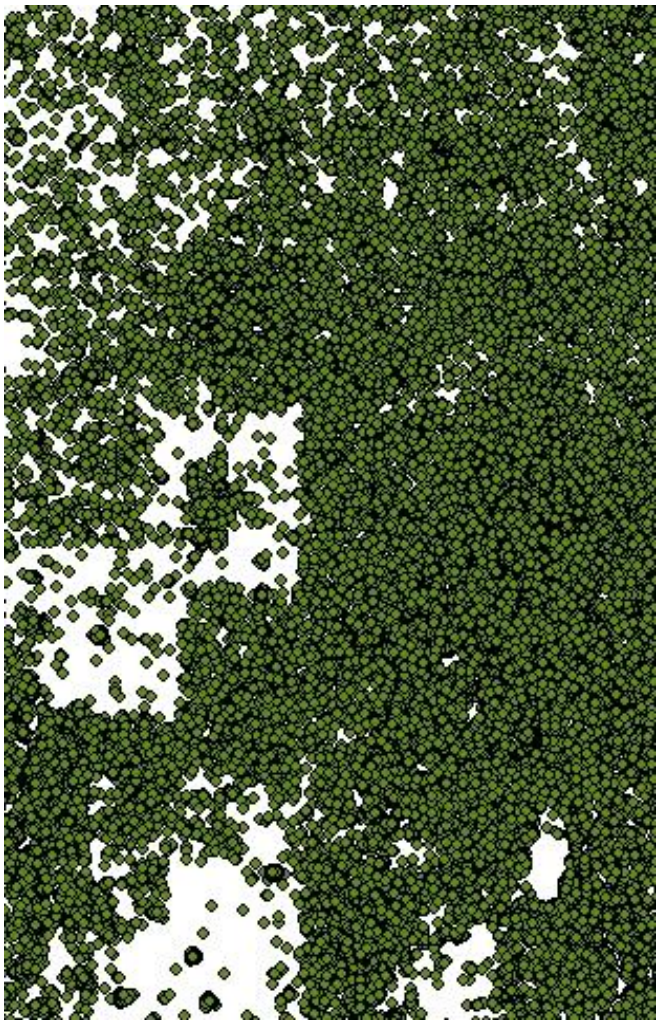


GIS in Health Begins with Geographic Coding

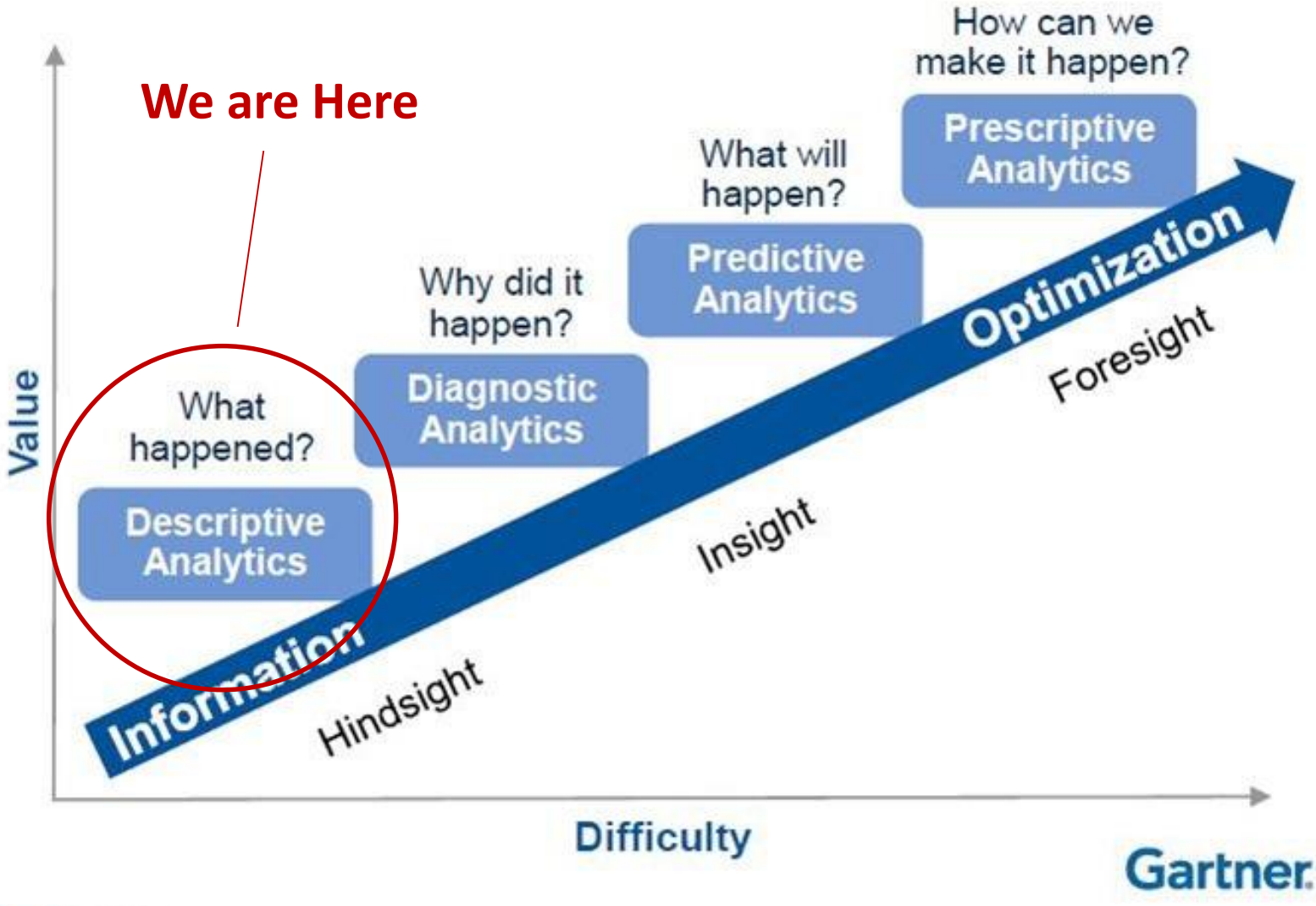
Records collected by ISDH

1999	C911	0	71	2.0000	1.0000	76
1999	C159	0	64	1.0000	1.0000	35
2000	E149	0	70	1.0000	1.0000	02
2000	J449	0	74	2.0000	2.0000	02
1999	J449	0	68	1.0000	2.0000	02
1999	D469	0	71	2.0000	1.0000	02
2000	J449	1	80	1.0000	1.0000	02
1999	I619	0	73	2.0000	1.0000	02
1999	E149	0	72	2.0000	1.0000	02
1999	J449	0	81	2.0000	2.0000	35
1999	I269	0	71	2.0000	1.0000	02
1999	I518	0	68	1.0000	1.0000	02
2000	I713	0	81	2.0000	2.0000	02
2000	E149	0	74	1.0000	1.0000	02
1999	V861	0	78	1.0000	1.0000	27
1999	I609	0	70	2.0000	2.0000	02
2000	I269	1	78	2.0000	1.0000	27
1999	J969	0	68	2.0000	2.0000	02
2000	J449	0	82	2.0000	2.0000	02
1999	E149	0	82	1.0000	1.0000	02
1999	I219	0	65	1.0000	1.0000	90
1999	I219	0	74	1.0000	2.0000	02
2000	I518	0	71	1.0000	2.0000	35
1999	I639	0	74	2.0000	2.0000	57
1999	I420	0	84	1.0000	2.0000	27
1999	I269	0	73	2.0000	1.0000	02
2000	J449	0	78	1.0000	2.0000	02
1999	I255	1	72	1.0000	1.0000	02
2000	I219	0	68	1.0000	1.0000	02
2000	E149	0	78	2.0000	2.0000	02
2000	I518	0	78	1.0000	1.0000	02
1999	R092	0	74	2.0000	1.0000	90
1999	D471	0	69	1.0000	1.0000	02
1999	C509	0	71	1.0000	1.0000	02
2000	I469	0	68	2.0000	2.0000	02
1999	J449	0	66	1.0000	1.0000	02
2000	E149	0	74	2.0000	1.0000	02

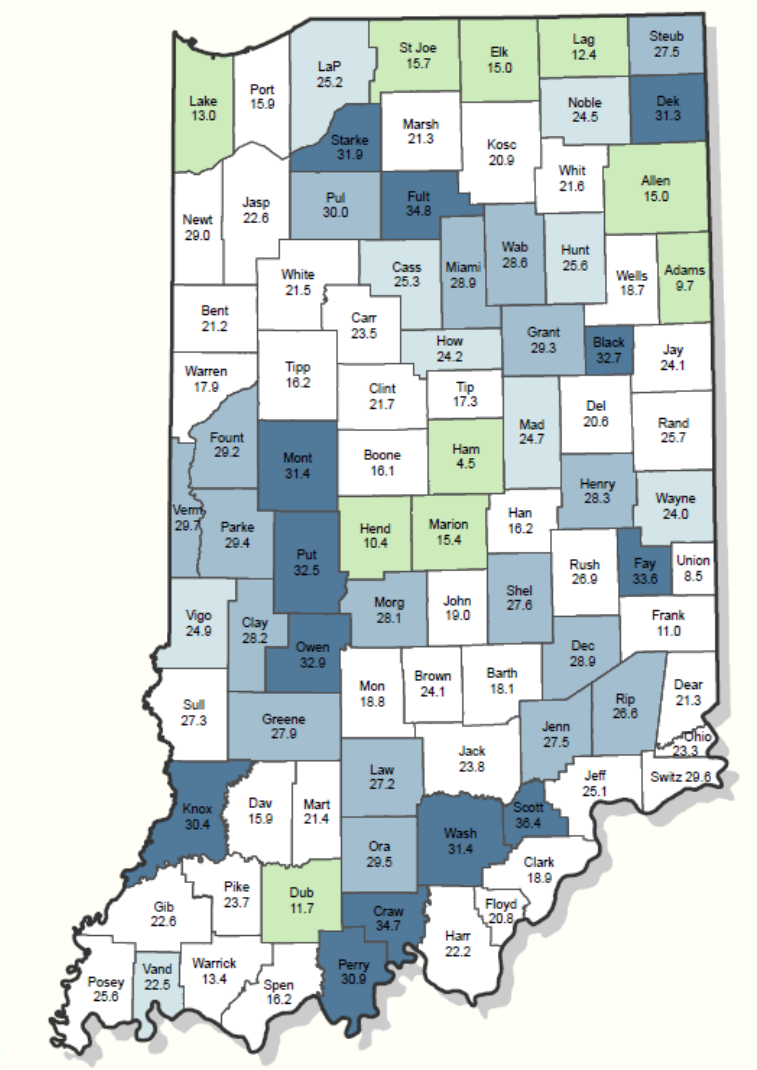
Records are a point on a map



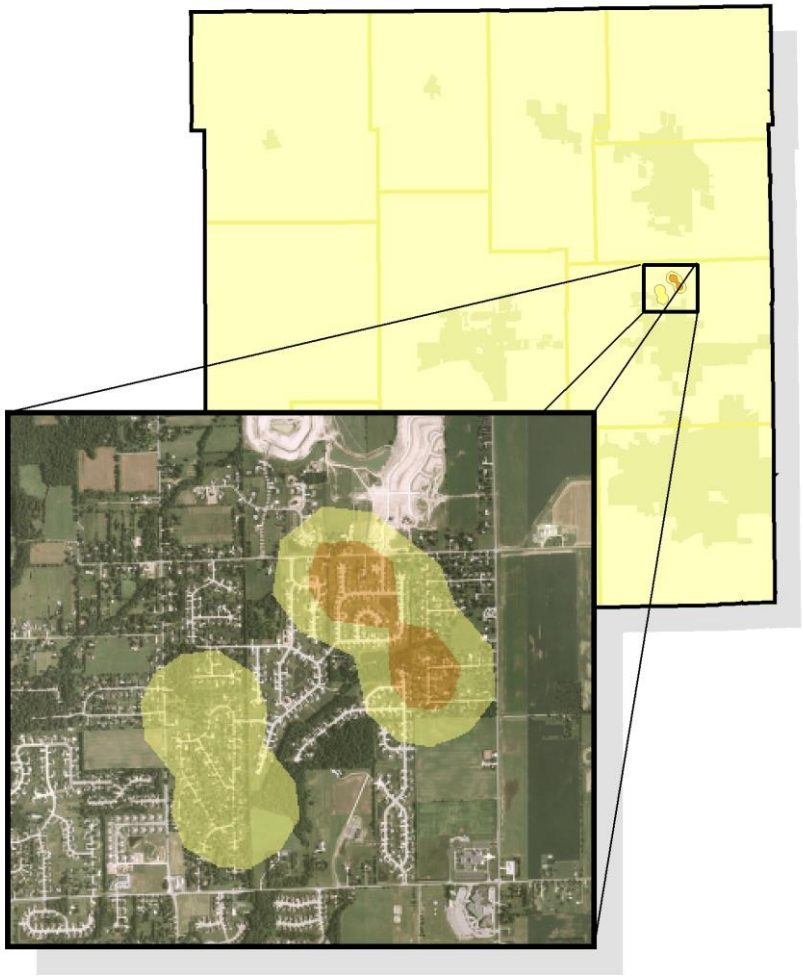
- **What we are repeatedly and increasingly asked for?**
  - Sub-county statistics
- **Can we share and distribute these stats?**
  - No, because of suppression rules (identifiability) and accuracy (noise) of stats due to less data available at such scales
- **How do we overcome these limitations?**
  - Provide statistics on **Observed** cases and events at increased geographic detail to focus health resources **Within** the county using a **Multi-scale Binning and Smoothing Methodology** that can be shared and distributed in the public domain to promote and support targeted health interventions


We are Here
How can we make it happen?
Prescriptive Analytics
What will happen?
Predictive Analytics
Why did it happen?
Diagnostic Analytics
What happened?
Descriptive Analytics
Value
Optimization
Foresight
Insight
Hindsight
Information
Difficulty
Gartner

Describing Health By County as Whole
(share data with everyone)



Targeting Health of Neighborhoods
(no sharing – actionable info not leveraged)



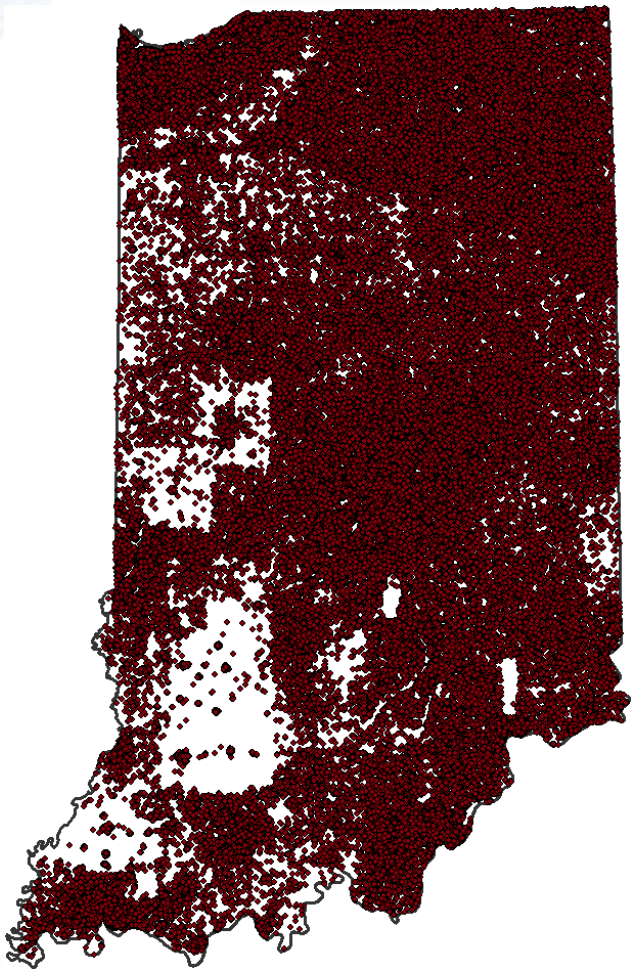
Finding the balance between coarse data and detailed data while maintaining data stability and confidentiality.

Not enough information for public/partners

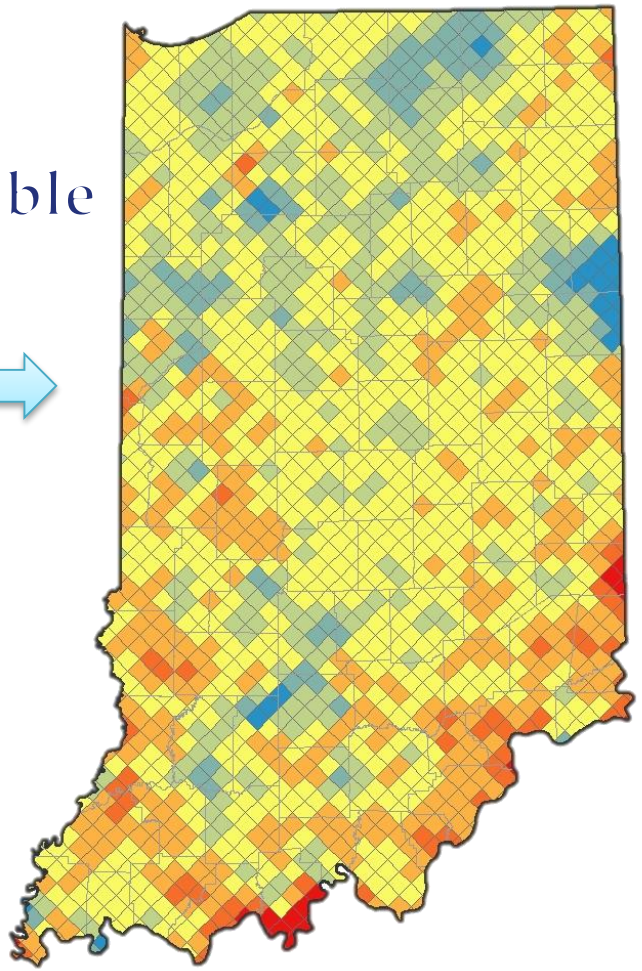
?

What is this solution?

Too much information for public/partners (identifiable)

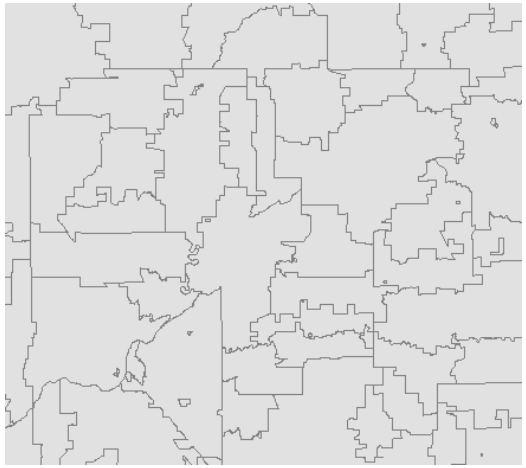


Data to Actionable Information



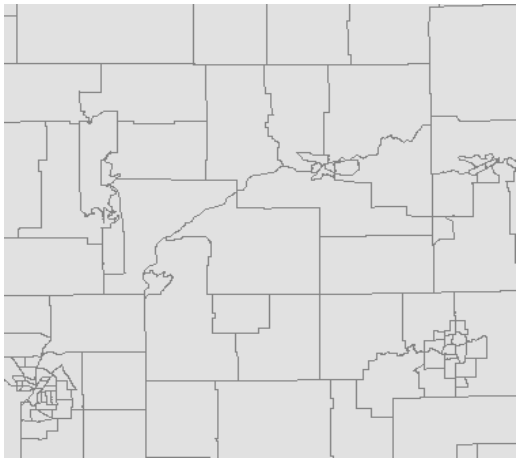
Common Sub-County Aggregation Areas

Zip Code



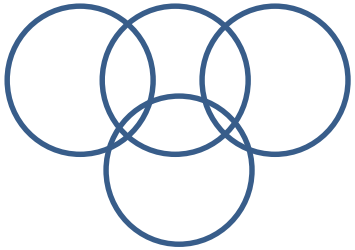
- Wide range of population (100 – 100,000)
- Point-based zips aren't often cross-walked to areas
- Small zips aren't often cross-walked to large zips
- Zip Codes do not exactly equal Census ZCTA
- Zip boundaries change
- Data collection doesn't check zip for accuracy

Census Tract



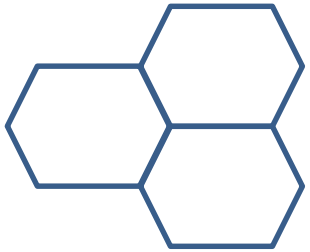
- Some tract populations are < 1,000
- Tracts can be very small areas
- Tracts can be oddly shaped
- Tracts boundaries change
- Tract geography is considered 'too identifiable'

Binning Options for Patient/Record Level Data



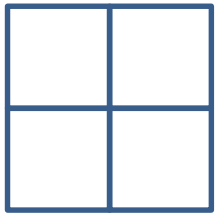
Used extensively by ISDH GIS in the past (Rushton)

Susceptible to false positives



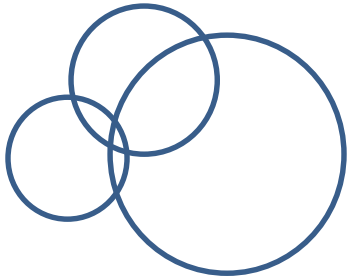
Currently a Popular option ("Hex-Binning")

Introduces directional bias



Straight-forward, Out-of-Box

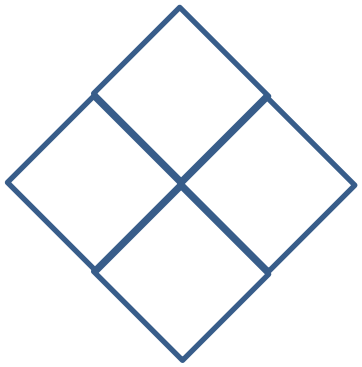
Arbitrary



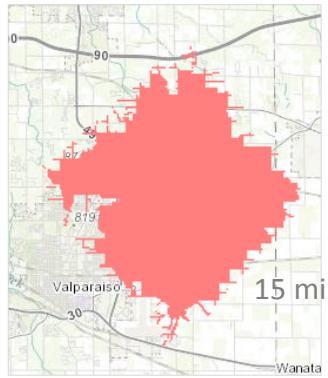
Ensures data stability (Rushton)

Large bins might not describe data at source point

Varying sizes of bins might be confusing



Diamond Binning – Based on Reasoning

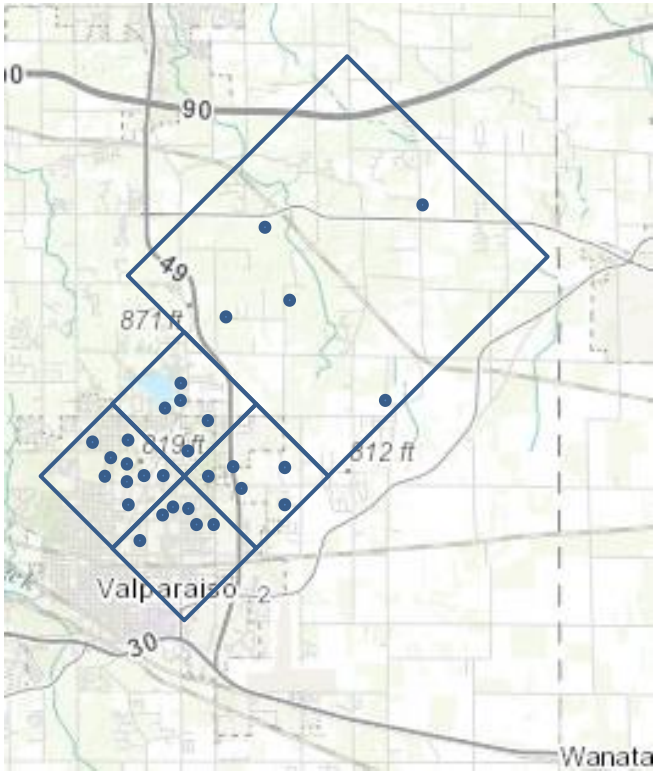


Drive-Time Service Area

The road grid system covers nearly all of Indiana. One can drive further when travelling north, south, east or west from a point than travelling NW, NE, SW or SE. The distance travelled for a given amount of time creates an extent boundary in a general shape of a diamond. Since neighborhoods and communities are closely tied to streets and people with tend to live near people of similar demographic characteristics, we reason that a diamond better captures a 'neighborhood' of people.

Concurrent Binning (Record Aggregation) for Urban and Rural Population

Accounting for a lot of data points AND too few data points

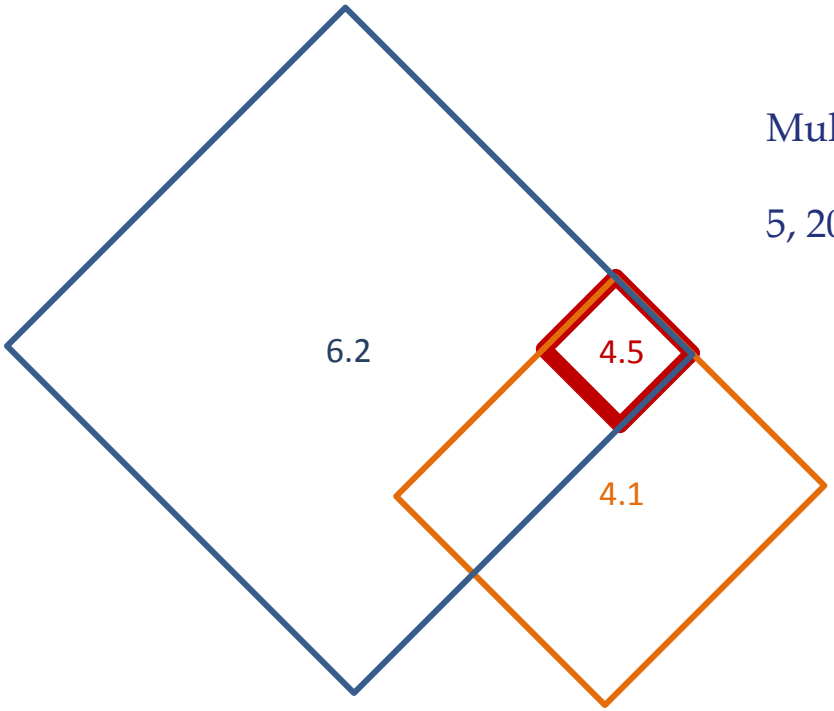


Multi-Scale:

Small bins for urban – more data points available in small area

Large bins for rural – more area needed to capture enough data points

Smoothing Bin Statistics to Eliminate Random Variability And Increase Data Reliability



Multi-Scale Offset Approach:

5, 20, and 80 square miles and N, S, E, and W of seed

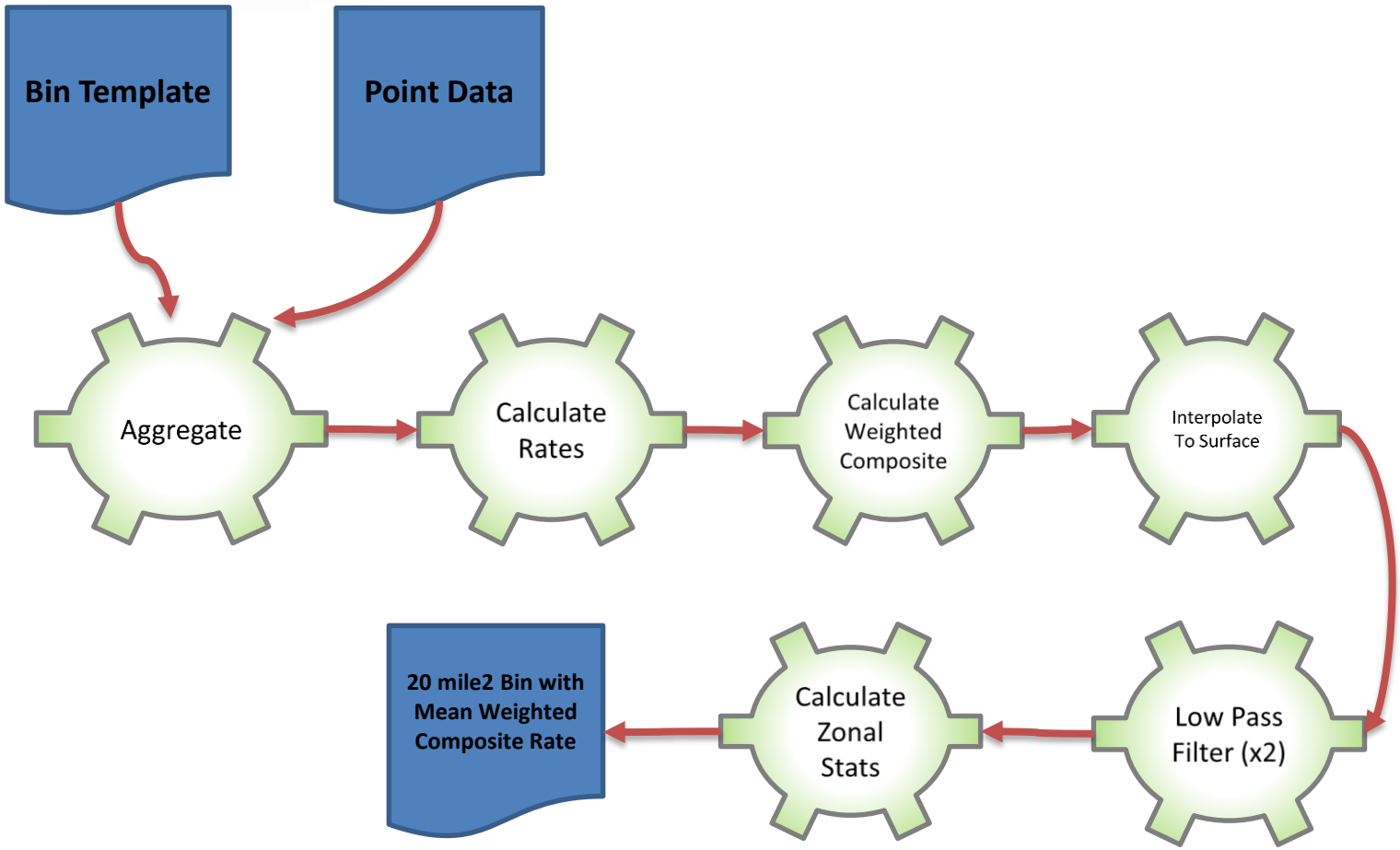
Python Inputs and Outputs

Inputs

- Point layer with risk factors
- Bin template
- Need to create ArcPro project ahead of time

Outputs

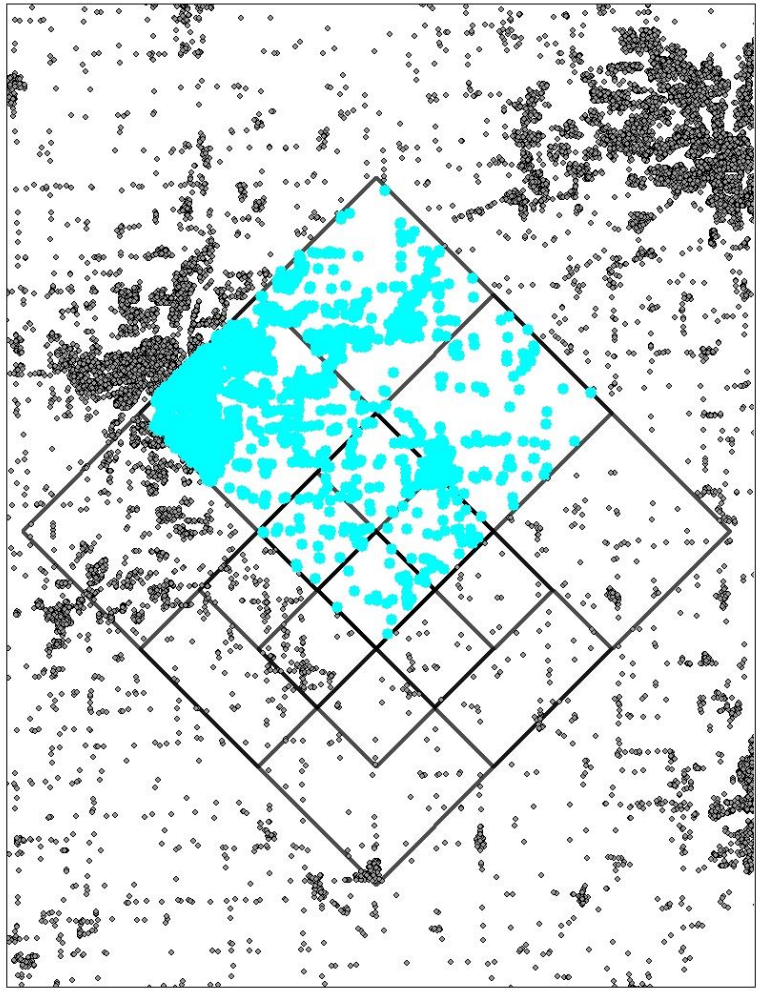
- 20 mile bins with mean weighted composite rate
- Point layer with significance for each diamond
- Intermediate bins and smoothed layers


Bin Template
Point Data
Aggregate
Calculate Rates
Calculate Weighted Composite
Interpolate To Surface
Low Pass Filter (x2)
Calculate Zonal Stats
20 mile2 Bin with Mean Weighted Composite Rate

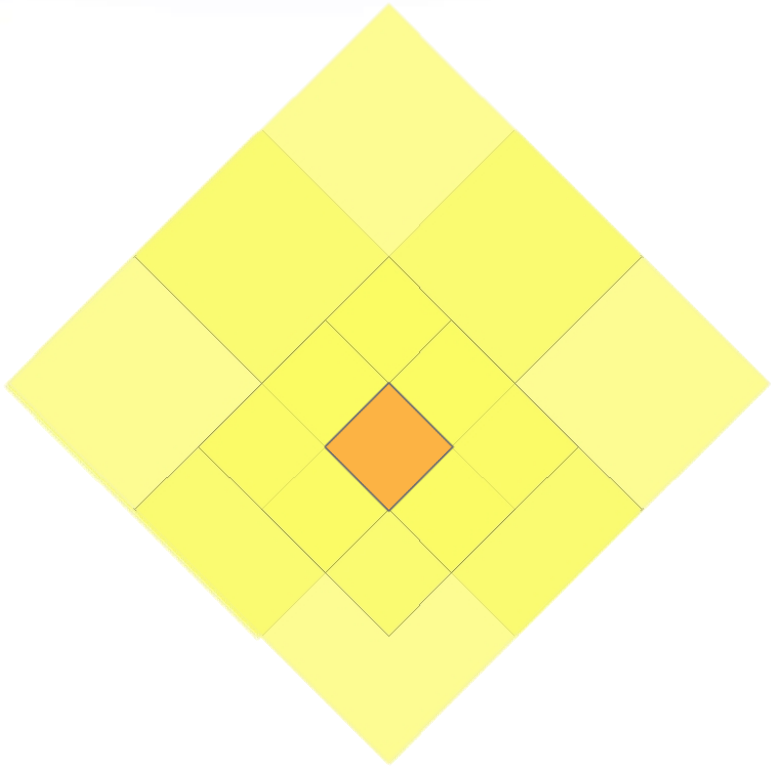
Multi-Scale Binning

Data is aggregated at 3 scales (5 mi^2, 20 mi^2, and 80 mi^2 *Diamonds*)

20 and 80 mile diamonds are shifted in 4 directions to reduce directional bias – Total of 9 multi-scale bins



Calculating Composite Rate



5 mi² Weighted Composite Rate* = ((3 * 5mi² Rate) + (2 * 20mi² Rate) + (2 * 20mi²_N Rate) + (2 * 20mi²_E Rate) + (2 * 20mi²_W Rate) + (80mi² Rate) + (80mi²_N Rate) + (80mi²_E Rate) + (80mi²_W Rate)) / 15

***Only bins with variable counts greater than user defined threshold (typically 20) are included in calculation**

5mi² Rate = 0.386364

20mi² Rate = 0.541295

20mi²_N Rate = 0.538126

20mi²_E Rate = 0.495726

20mi²_W Rate = 0.534035

80mi² Rate = 0.537832

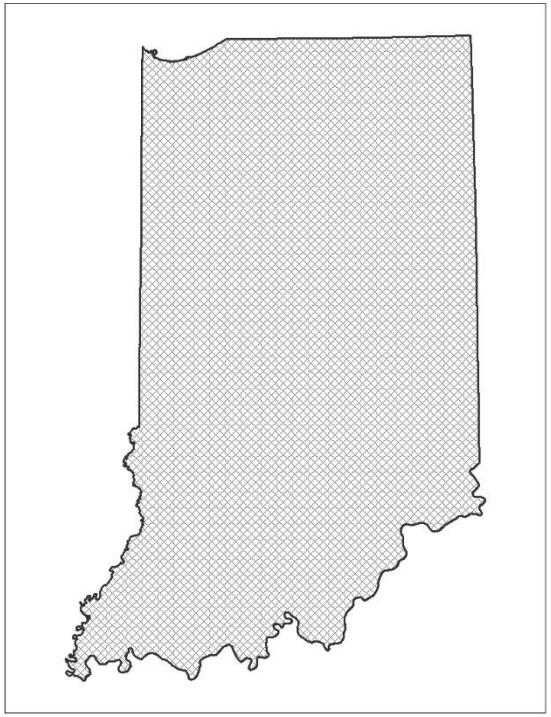
80mi²_N Rate = 0.536505

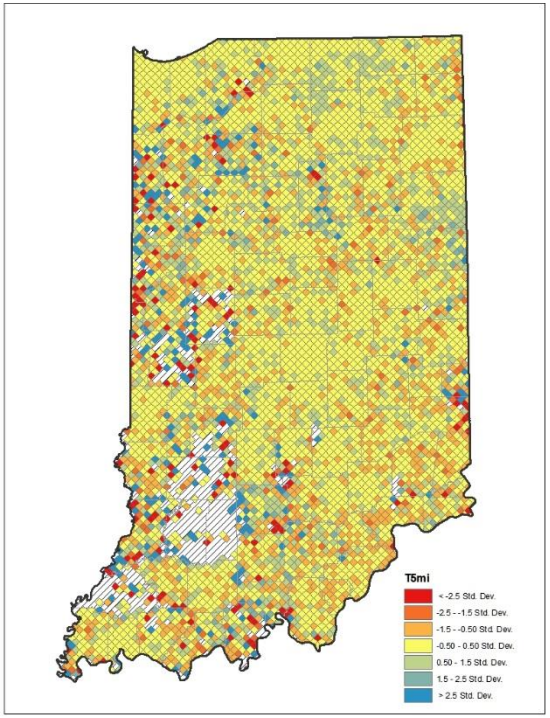
80mi²_E Rate = 0.536672

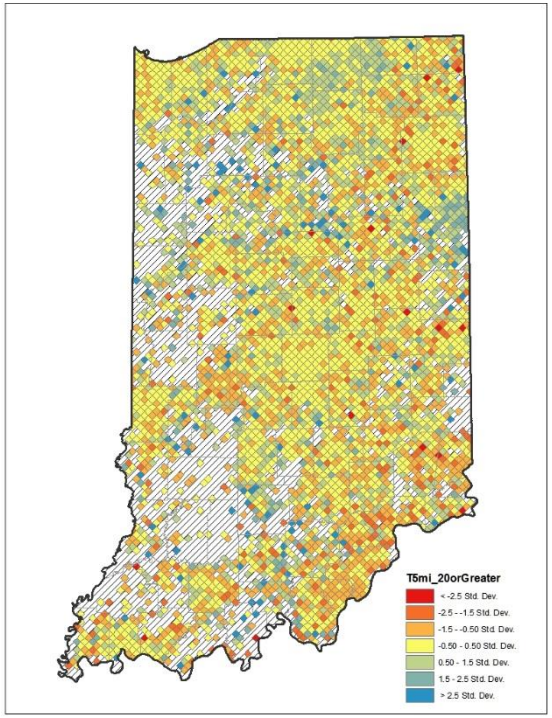
80mi²_W Rate = 0.536396

Weighting ensures that local data is more important in calculation of composite rate

0.501657 = ((3 * 0.386364) + (2 * 0.541295) + (2 * 0.538126) + (2 * 0.495726) + (2 * 0.534035) + (0.537832) + (0.536505) + (0.536672) + (0.536396)) / 15



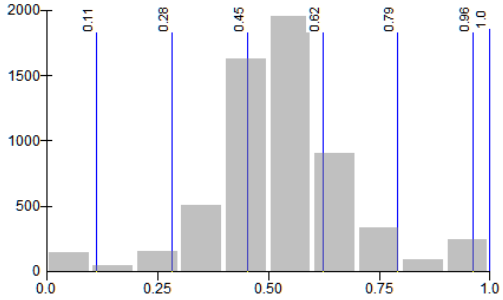


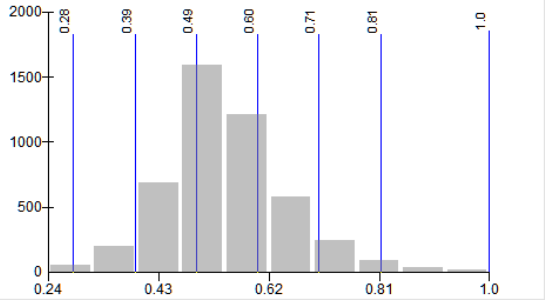


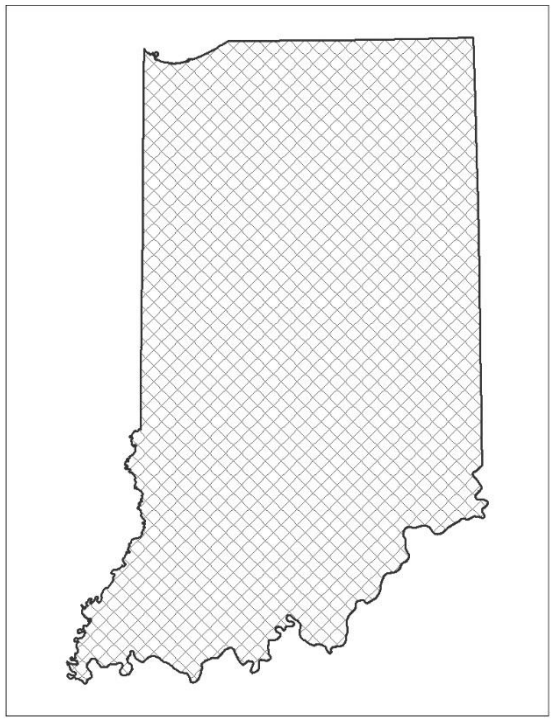
5 mi²

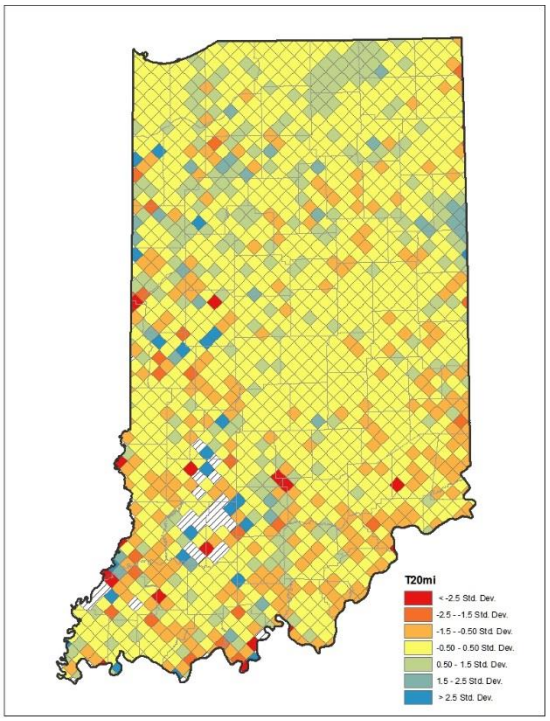
"seed"

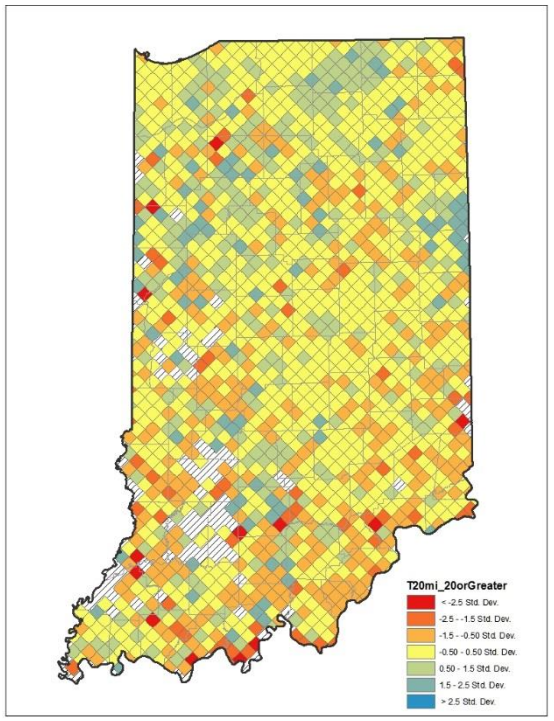
*** Histogram scale may not match.





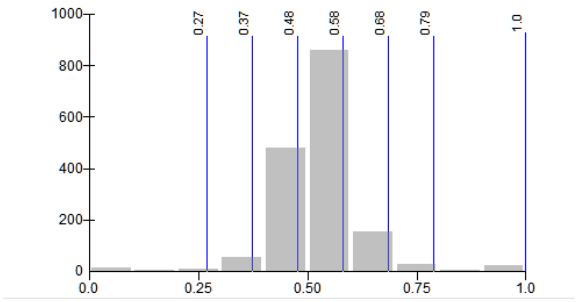


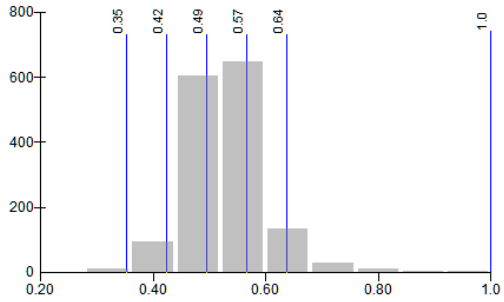


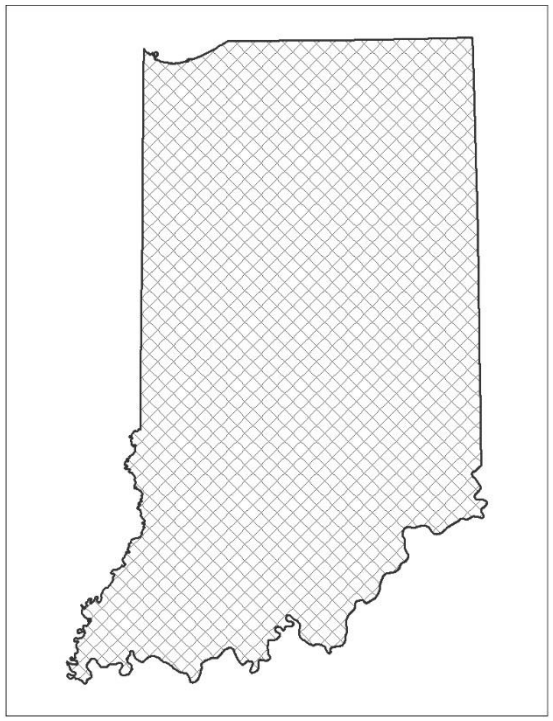


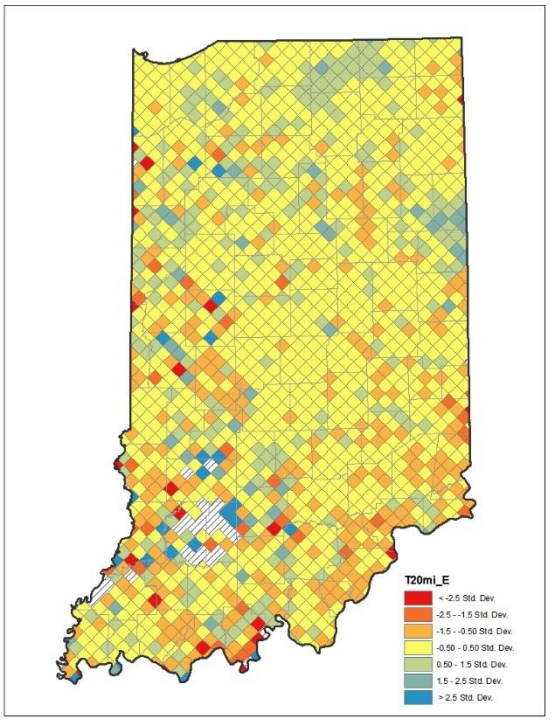
20 mi^2

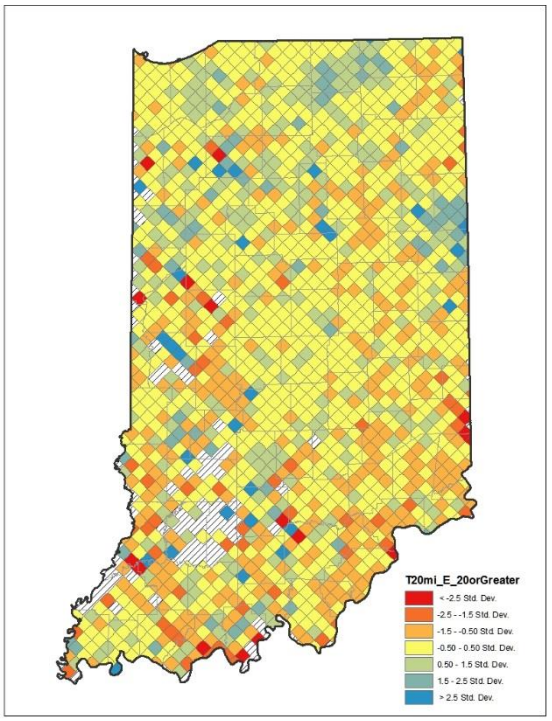
*** Histogram scale may not match.





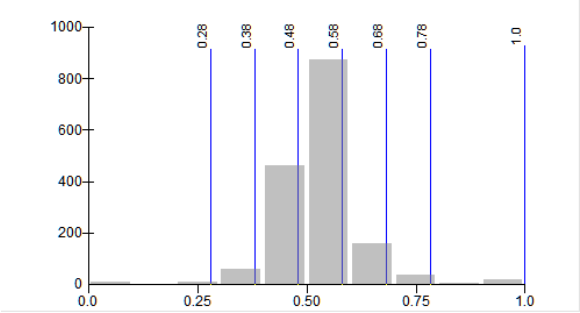


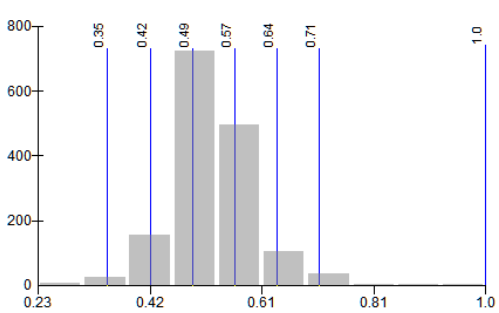


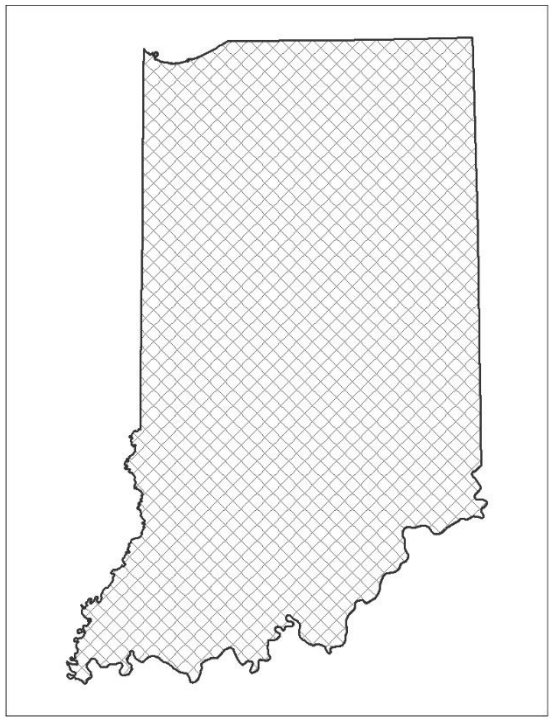


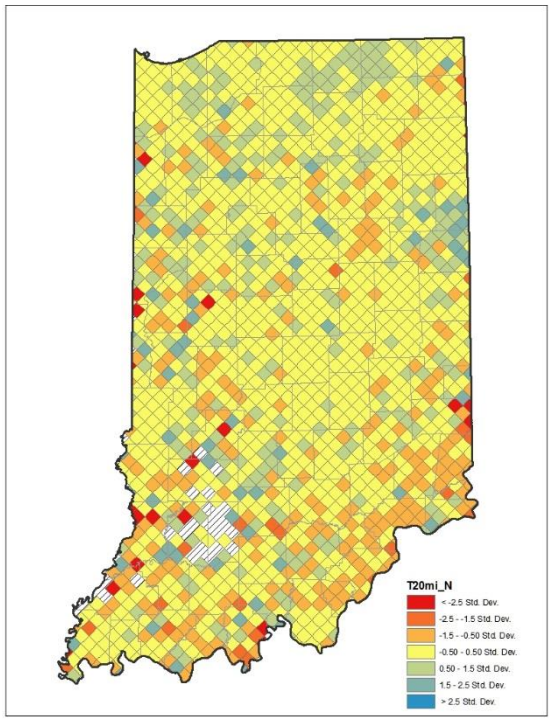
20 mi² East

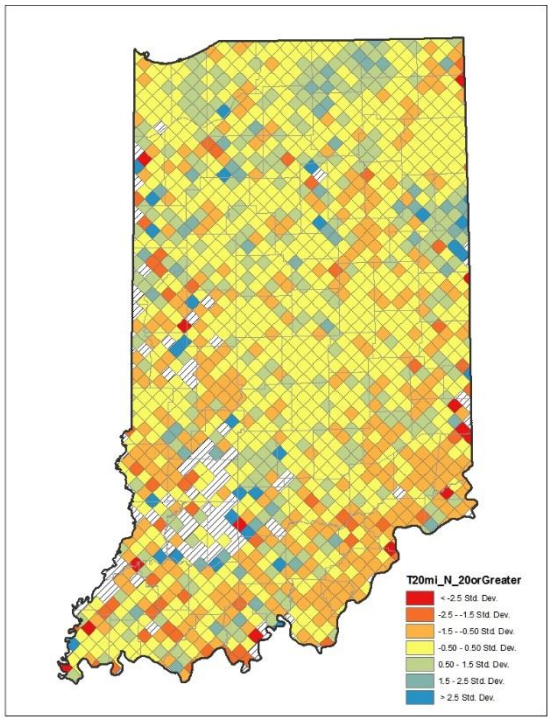
*** Histogram scale may not match.





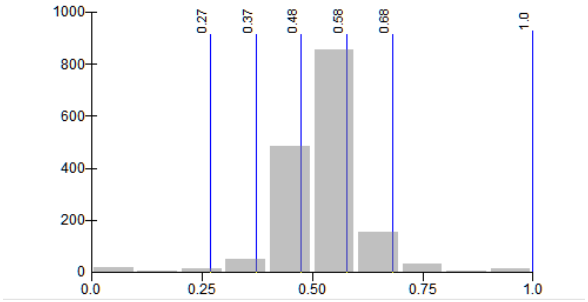


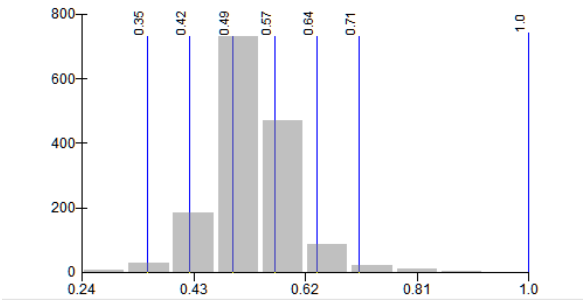


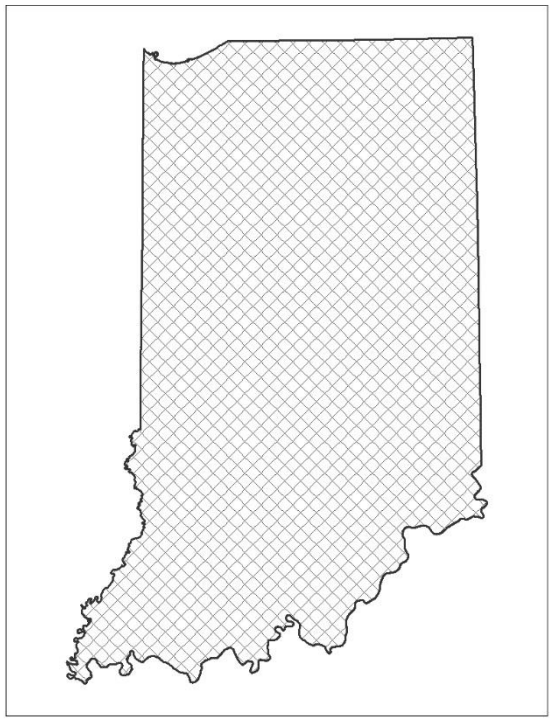


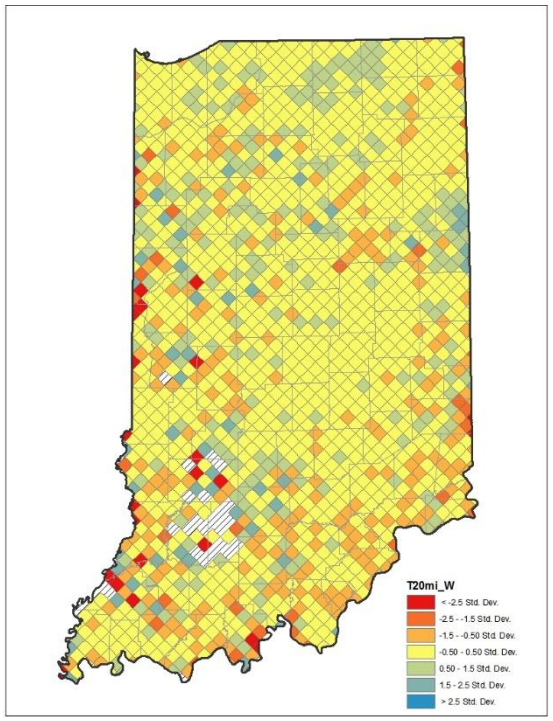
20 mi² North

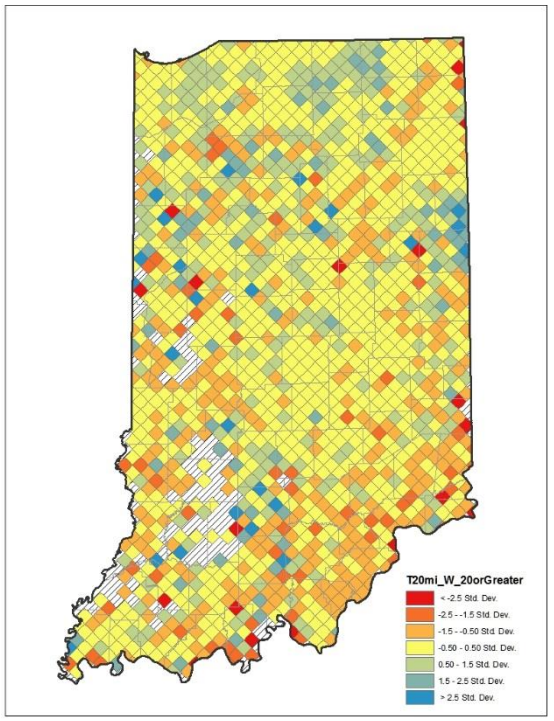
*** Histogram scale may not match.





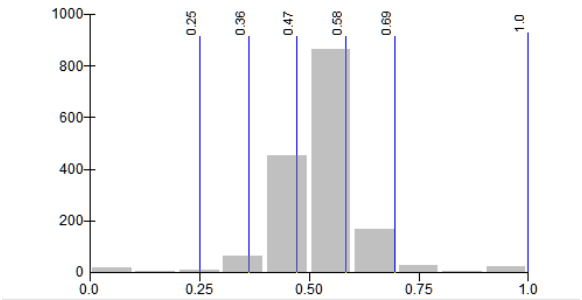


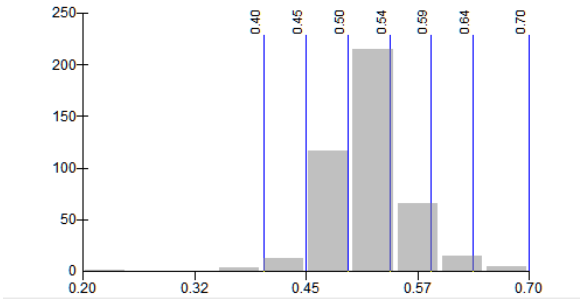


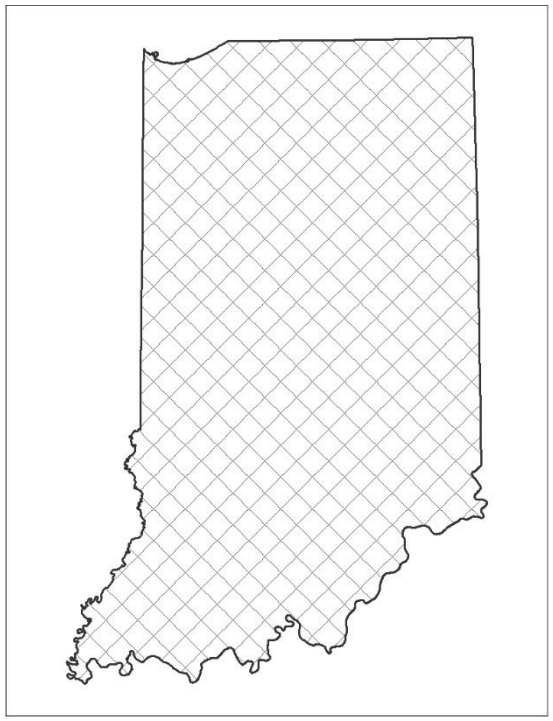


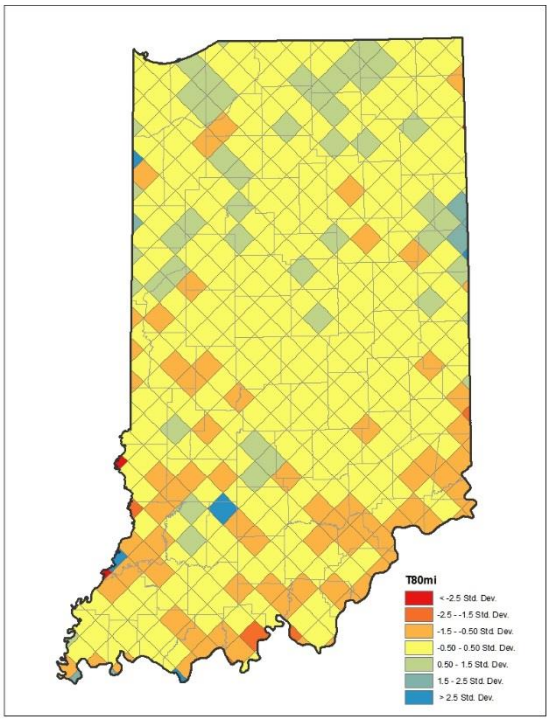
20 mi² West

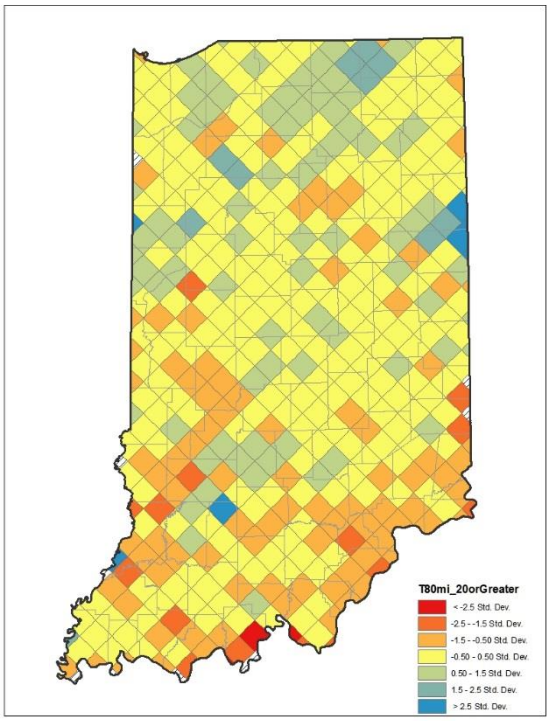
*** Histogram scale may not match.



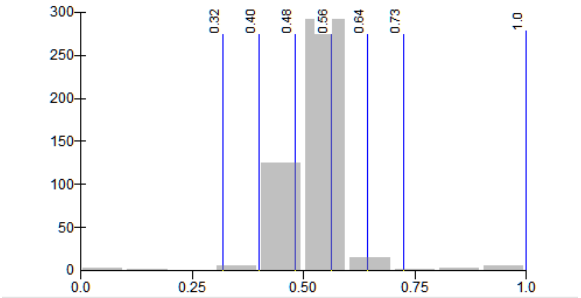


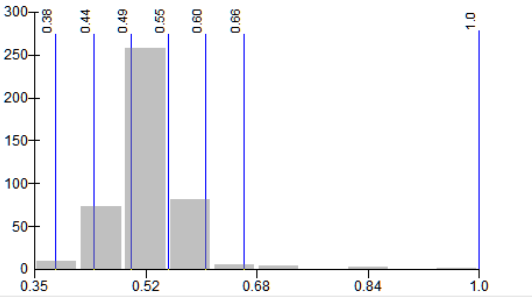




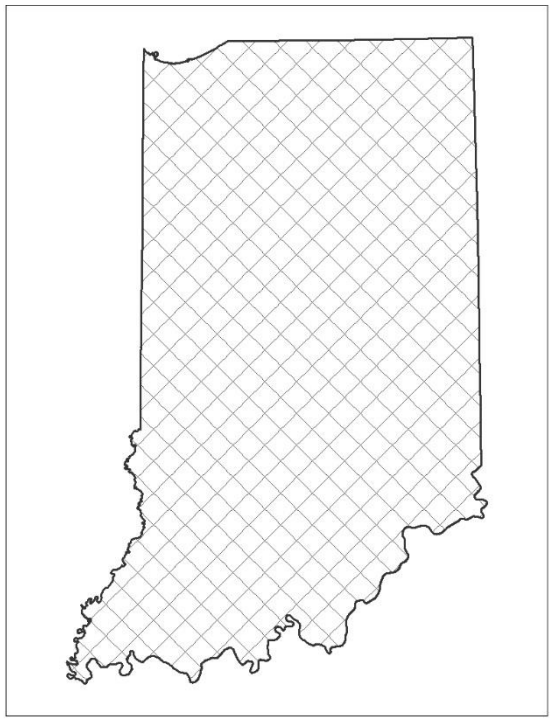


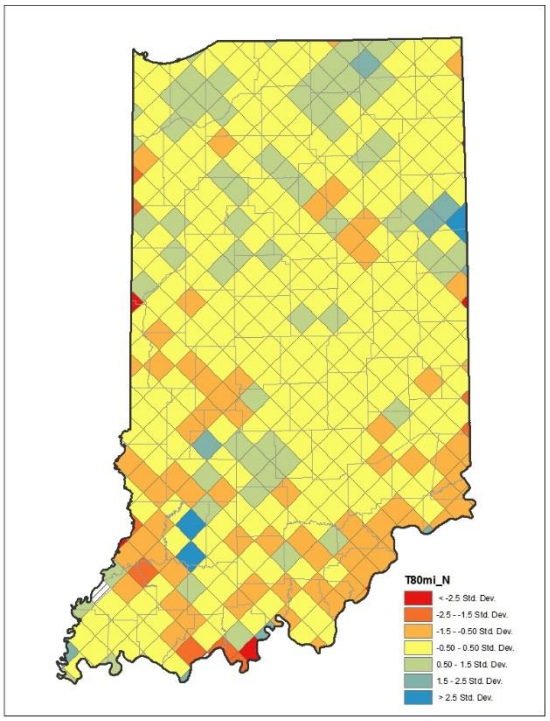
80 mi^2

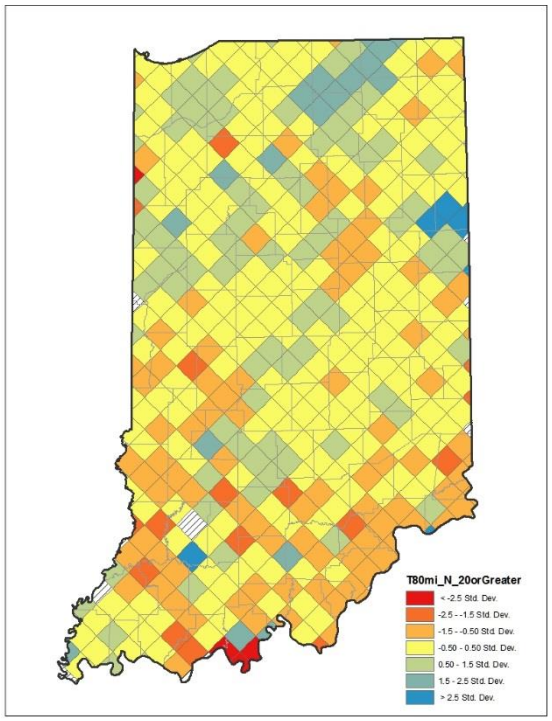




*** Histogram scale may not match.

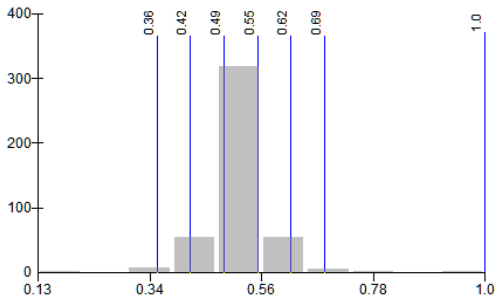


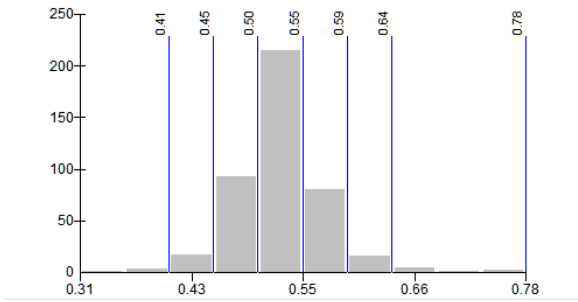


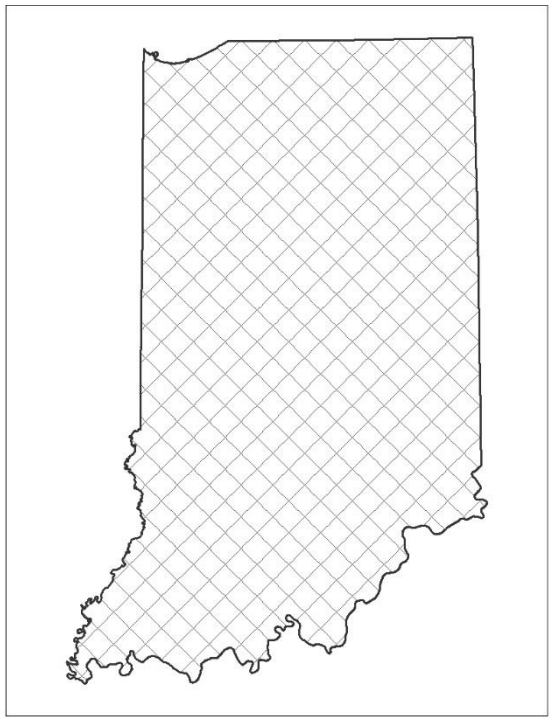


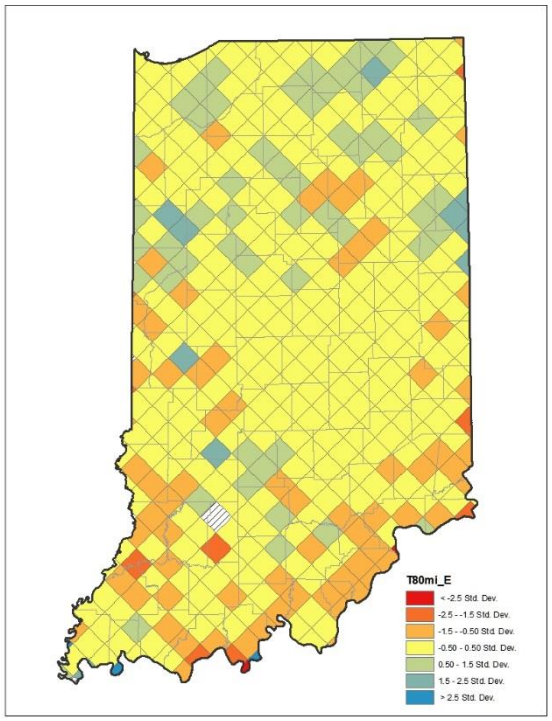
80 mi² North

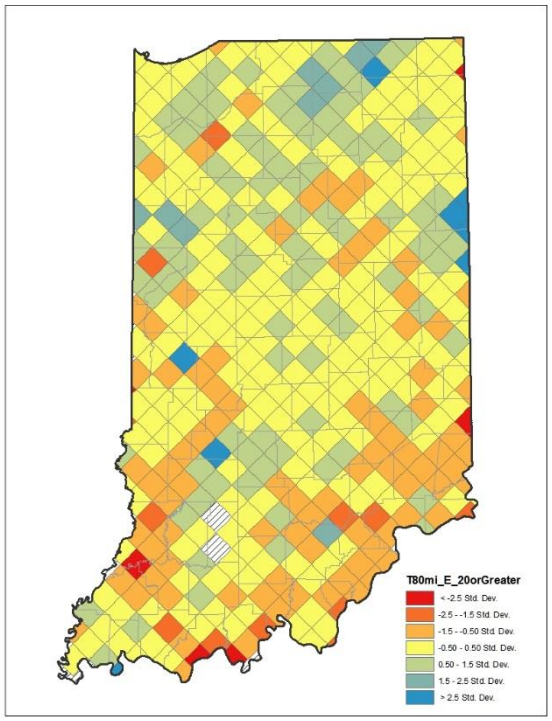
*** Histogram scale may not match.





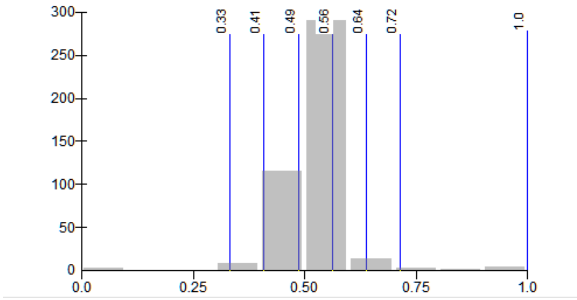


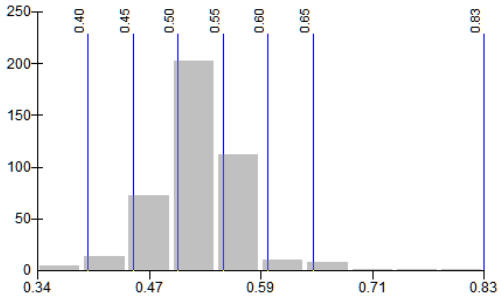


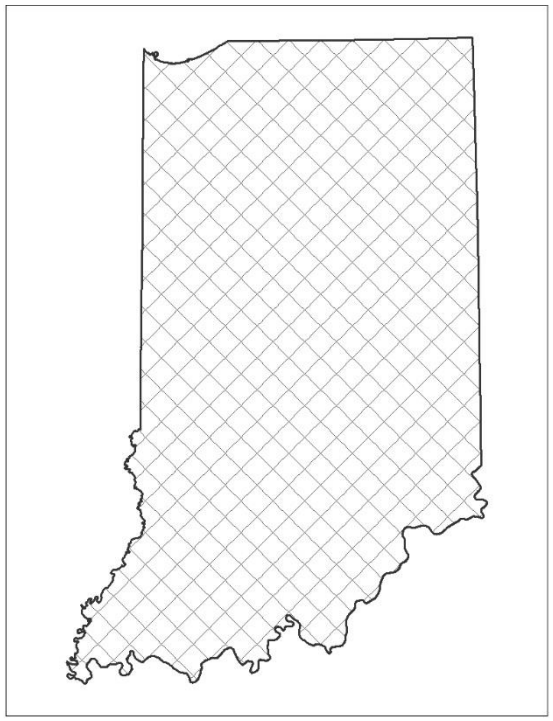


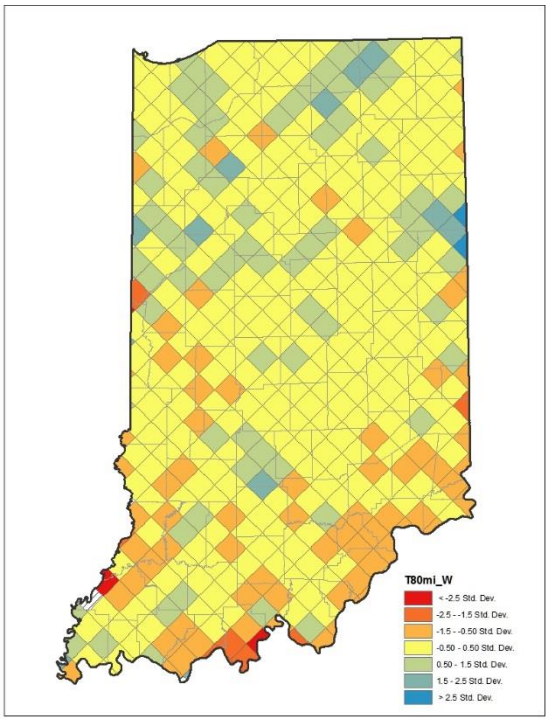
80 mi² East

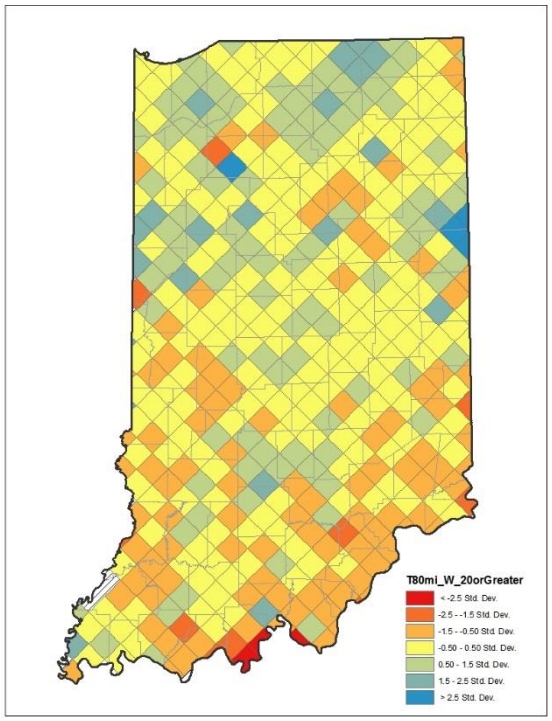
*** Histogram scale may not match.





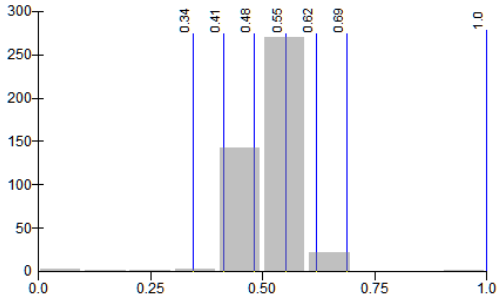


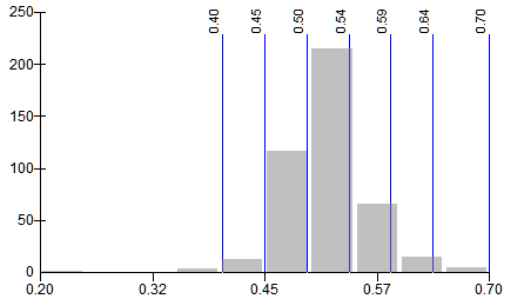




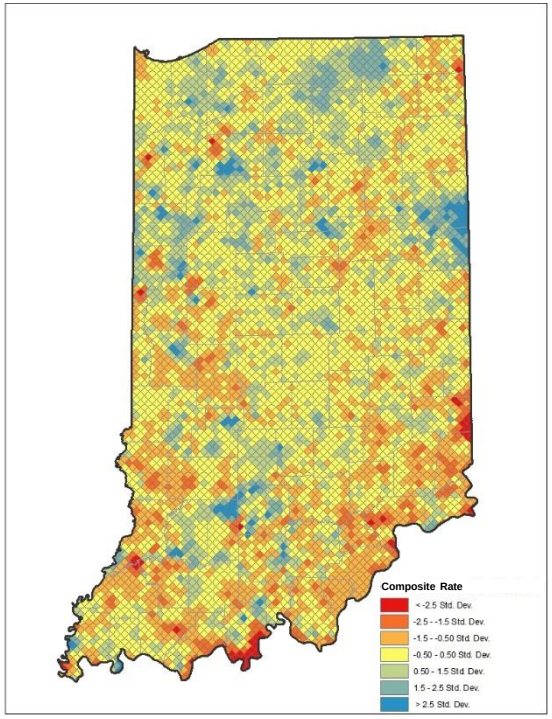
80 mi² West

*** Histogram scale may not match.





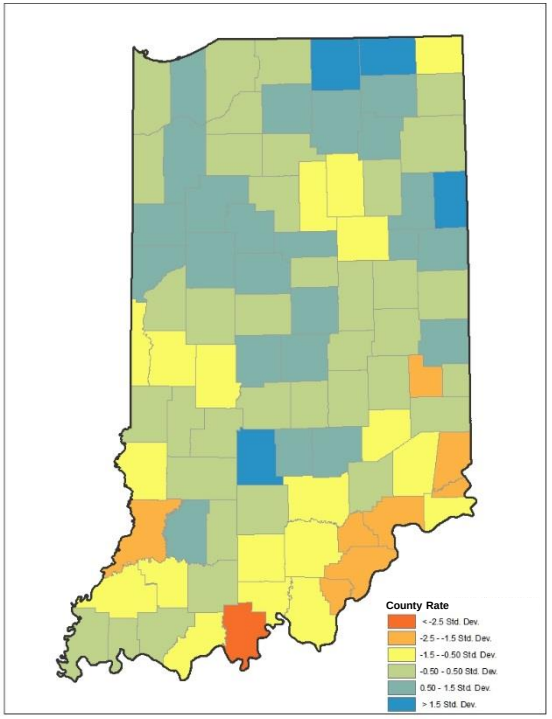
Initial 5mi² Composite Result

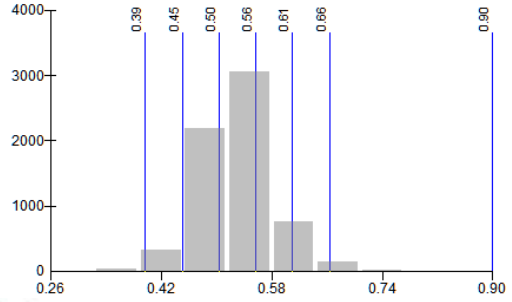


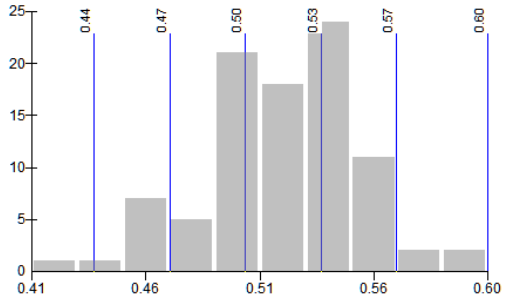
5 mile weighted composite rate contains more detail than county rates

but

May be too much detail to detect state-wide pattern

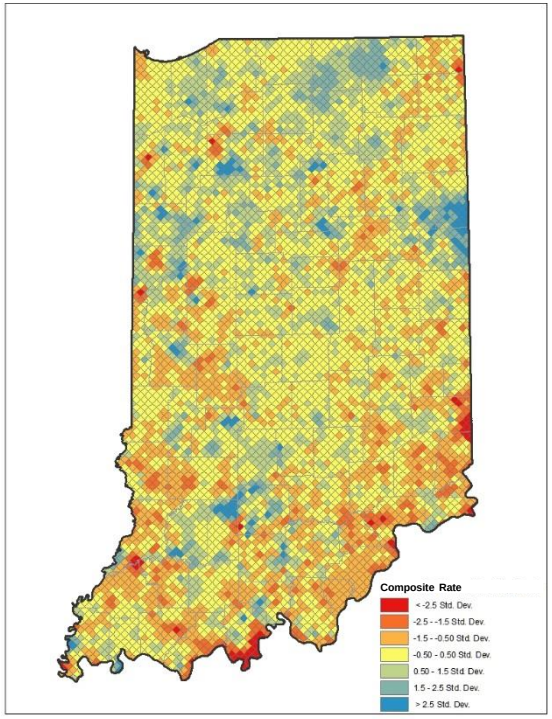


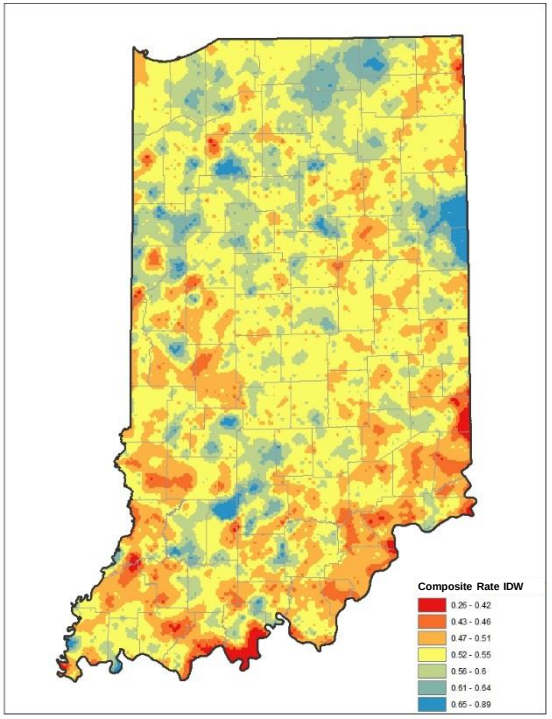


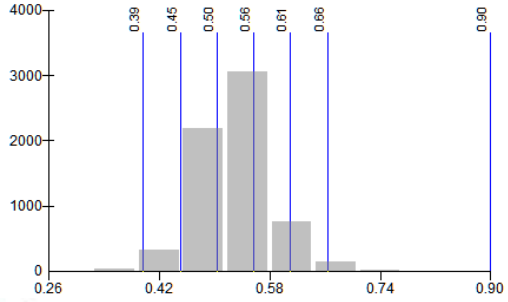


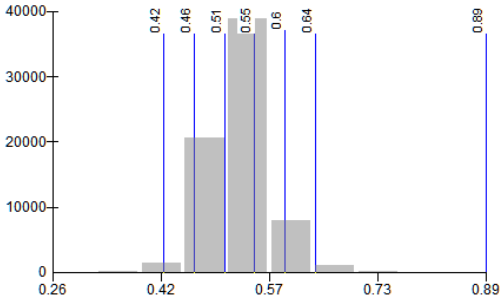
*** Histogram scale may not match.

Interpolate from Discrete Bins to Surface



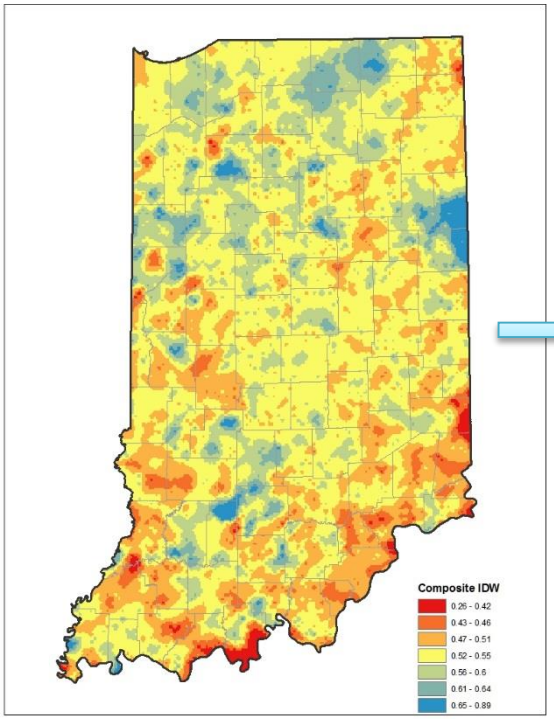


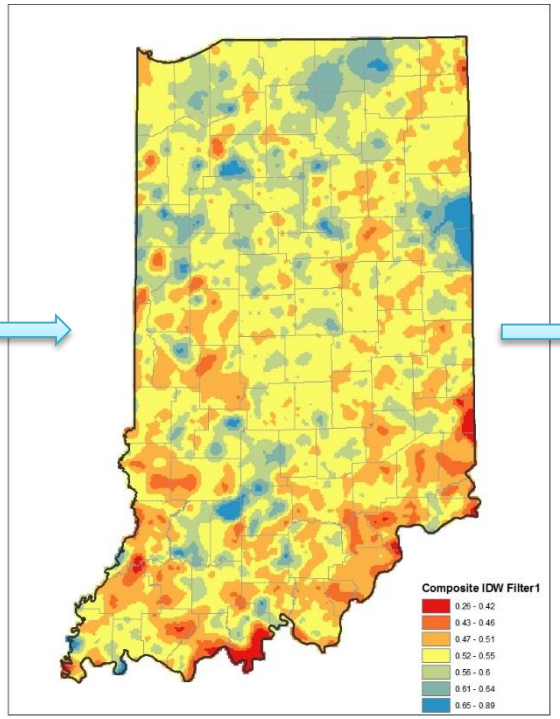


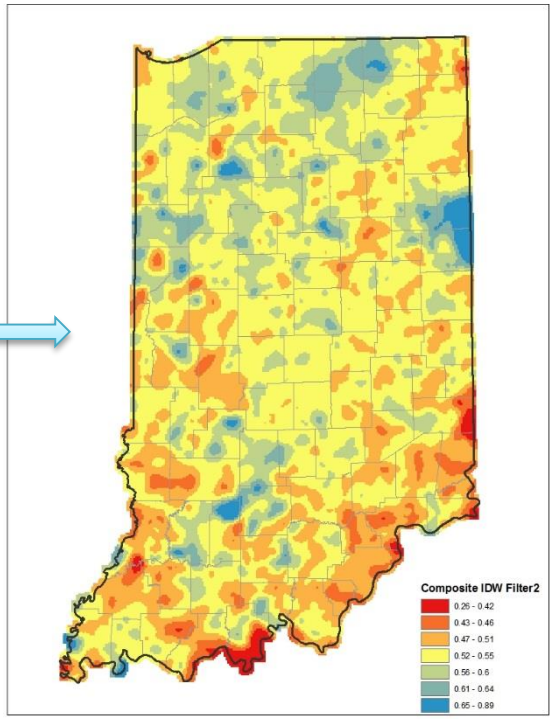


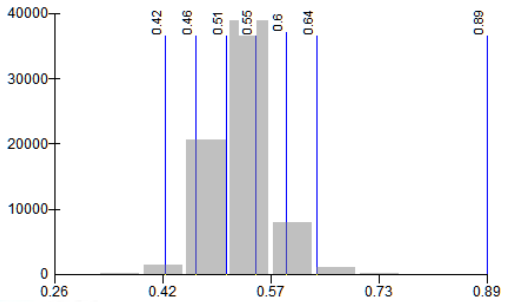
*** Histogram scale may not match.

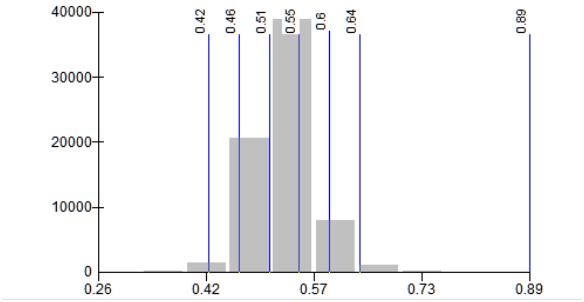
Raster Surface is Smoothed Using a Low Pass Filter Method to Remove Noise

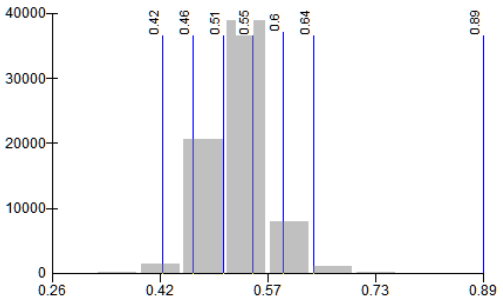




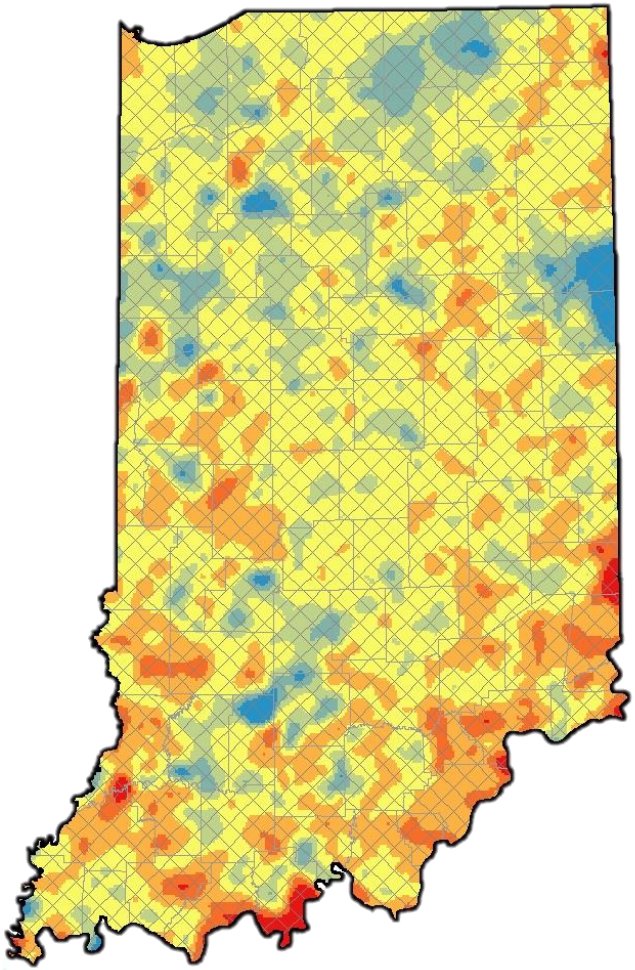






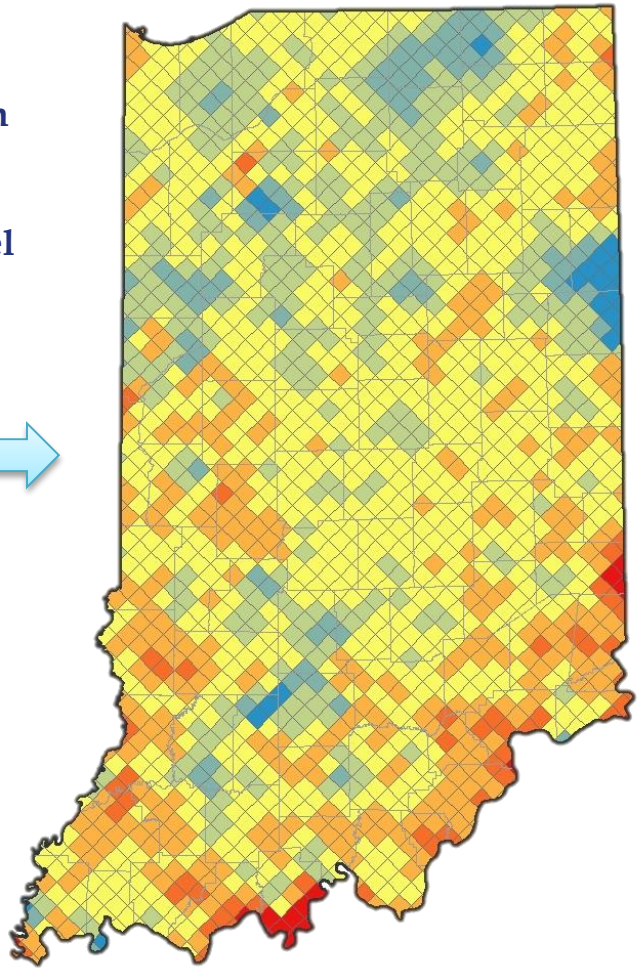


Calculate 20 mi2 Zonal Statistics



Surface statistics are calculated within each 20mi` bin

The mean of each pixel value represents the smoothed weighted composite rate

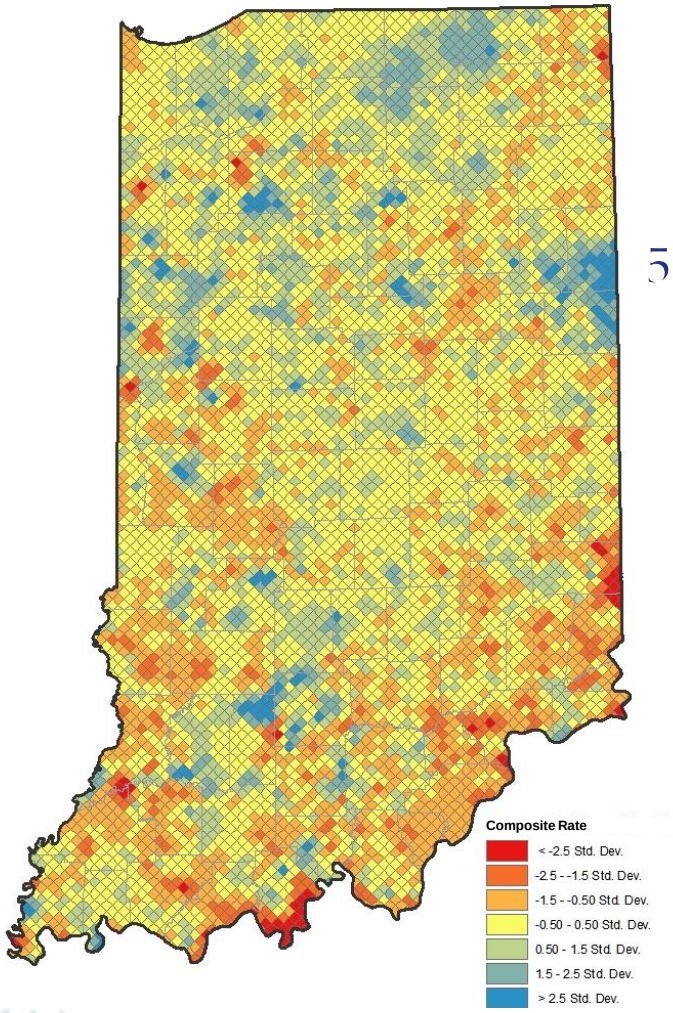


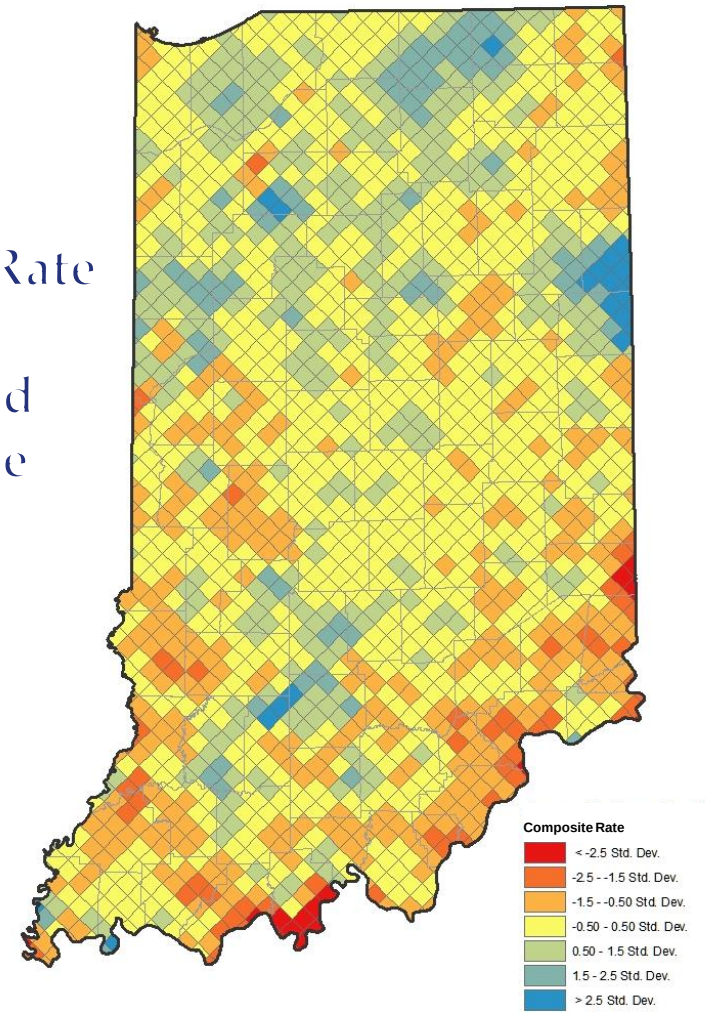
Compare:

5 mi^2 Composite Rate

vs

20mi^2 Smoothed Composite Rate





Link to web app containing GUMSS maps

4 Things to Remember When Interpreting the Maps

1

The value of a diamond bin is based on data within and around that diamond bin
(bins are smoothed)

2

Values of the diamond are based on patient records that could be geocoded
(typical geocode percentages are about 90% statewide)

3

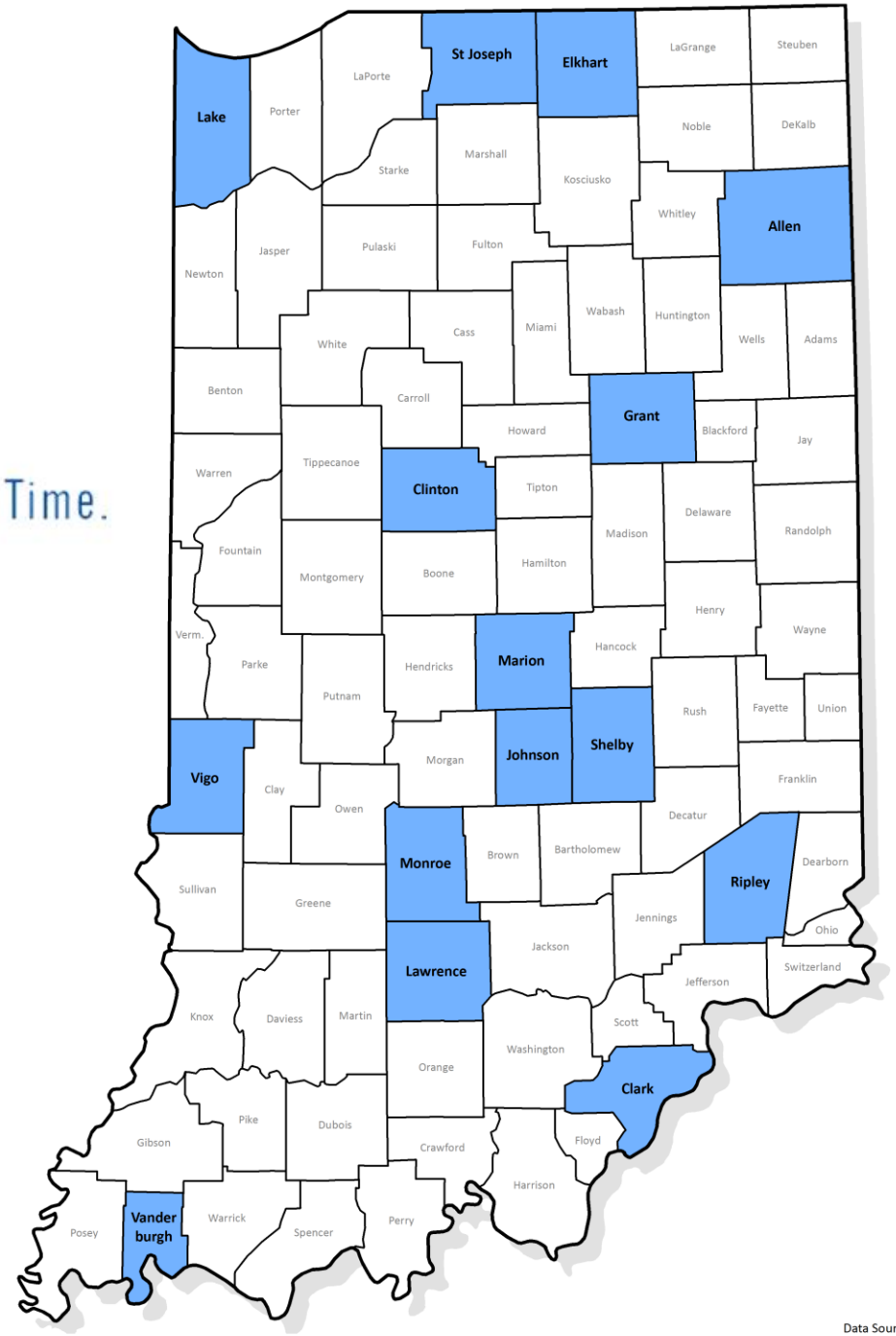
Values in diamonds with zero or a small number of geocoded points
rely more on data further away
(interpolation or inference to fill data gaps)

4

The process is built on the assumption of spatial autocorrelation
(like values tend to be nearer to one another in space)

Baby & me™
Tobacco Free

Healthy Babies. Born on Time.



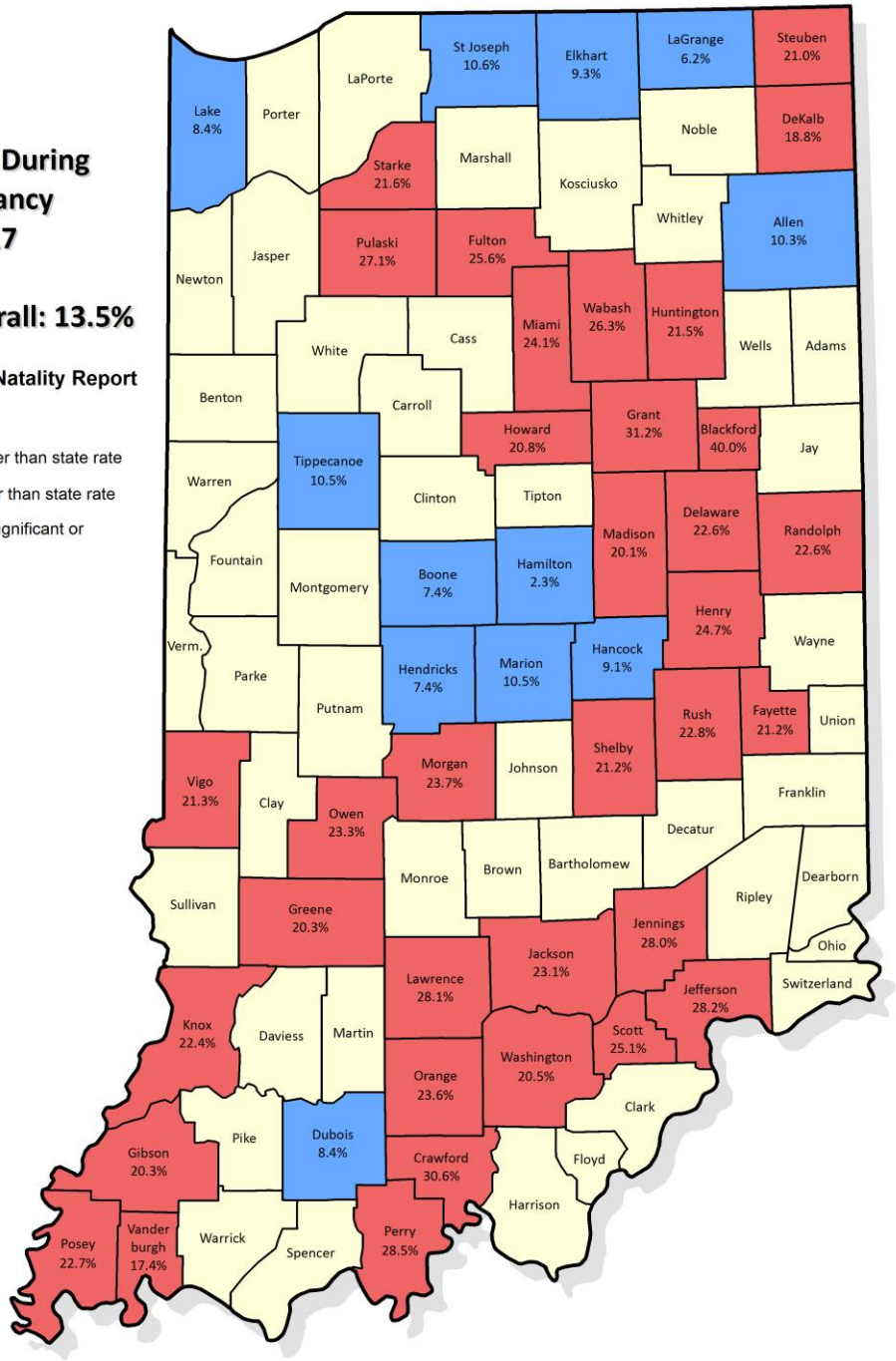
Data Source: ISDH MCH

Smoking During Pregnancy 2017

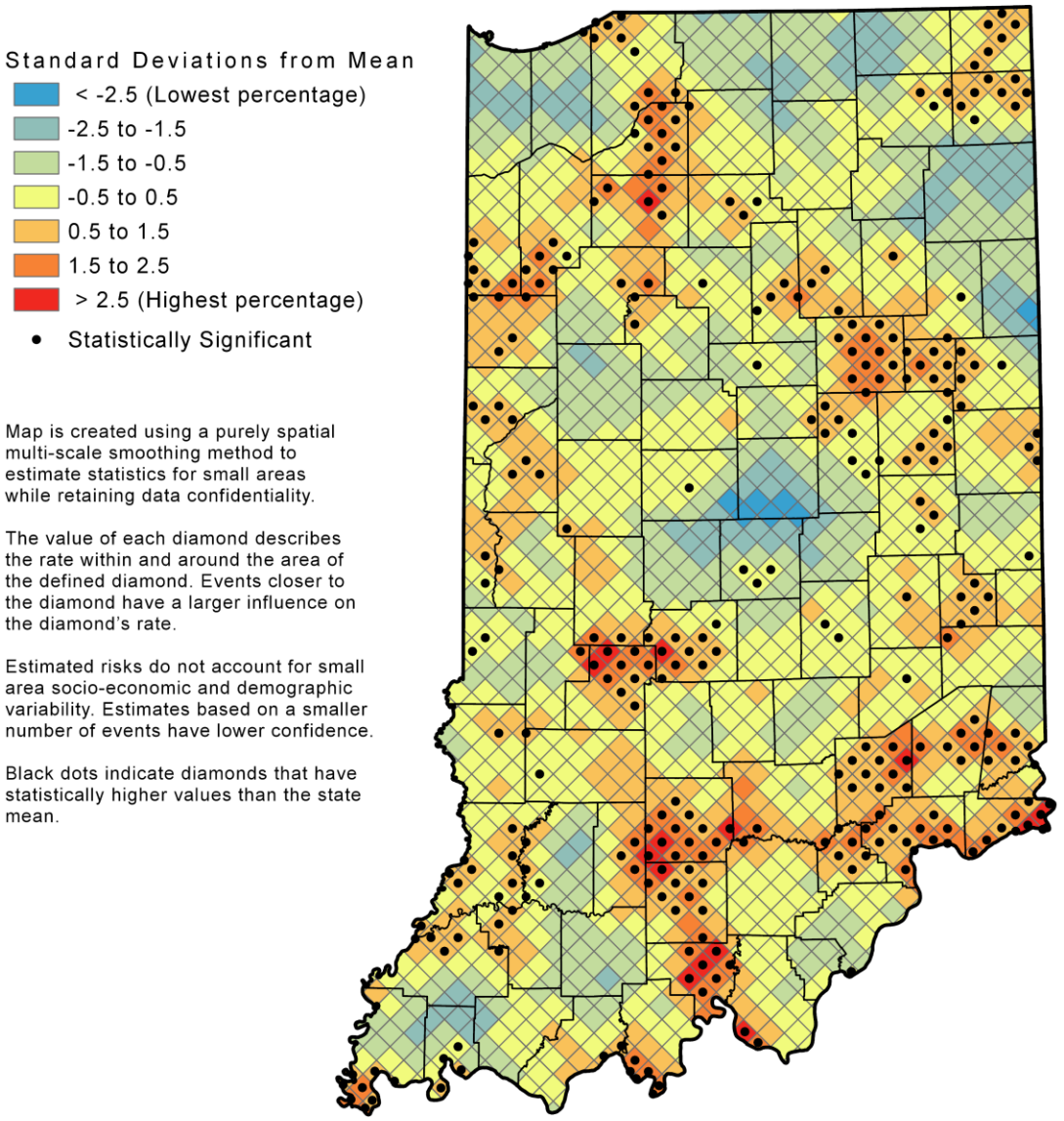
Indiana Overall: 13.5%

Source: Indiana Natality Report

- Significantly higher than state rate
- Significantly lower than state rate
- Not staitistically significant or unstable rate



Percent of Live Births Mother Smoked During Pregnancy, 2013 – 2017



Map is created using a purely spatial multi-scale smoothing method to estimate statistics for small areas while retaining data confidentiality.

The value of each diamond describes the rate within and around the area of the defined diamond. Events closer to the diamond have a larger influence on the diamond's rate.

Estimated risks do not account for small area socio-economic and demographic variability. Estimates based on a smaller number of events have lower confidence.

Black dots indicate diamonds that have statistically higher values than the state mean.

Map Author: ISDH ERC PHG 7/2019

Data Source: ISDH PHG, Data Analysis Team, Vital Records

Map Utilization

- Maternal & Child Health Needs Assessment research and outreach to high risk areas
- Grant proposals and program funding
- Maternal and Child Health strategic planning
- Education and data dissemination

Questions?

Robert Gottlieb
RGottlieb@isdh.IN.gov

Jenny Durica
JDurica@isdh.IN.gov