

Spatial Clustering and Autocorrelation

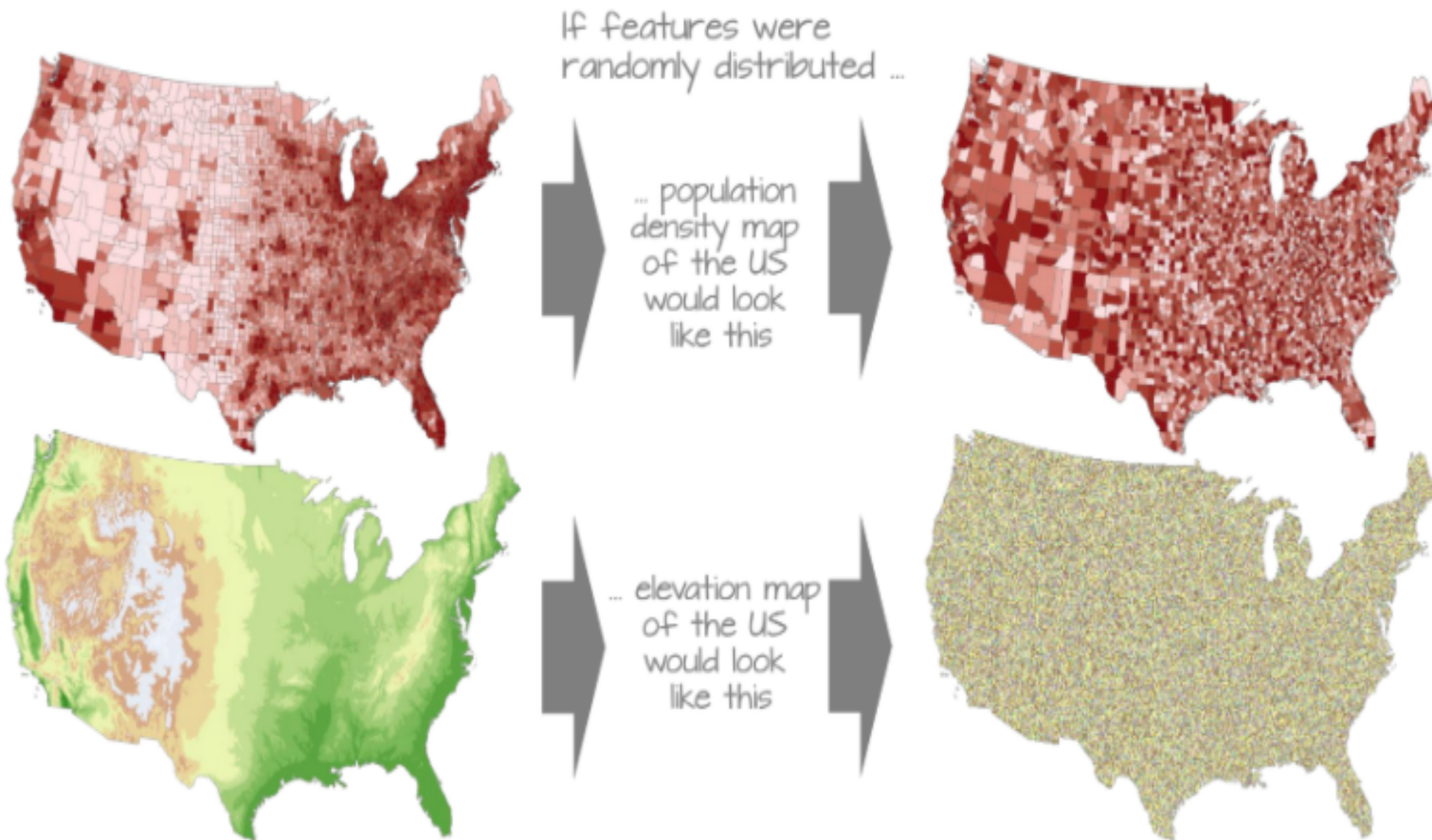
Class #20 | GEOG 215
Introduction to Spatial Data Science
Varun Goel

Today's Class

- Spatial Cluster Analysis
- Spatial Pattern
- Autocorrelation

Tobler's First Law

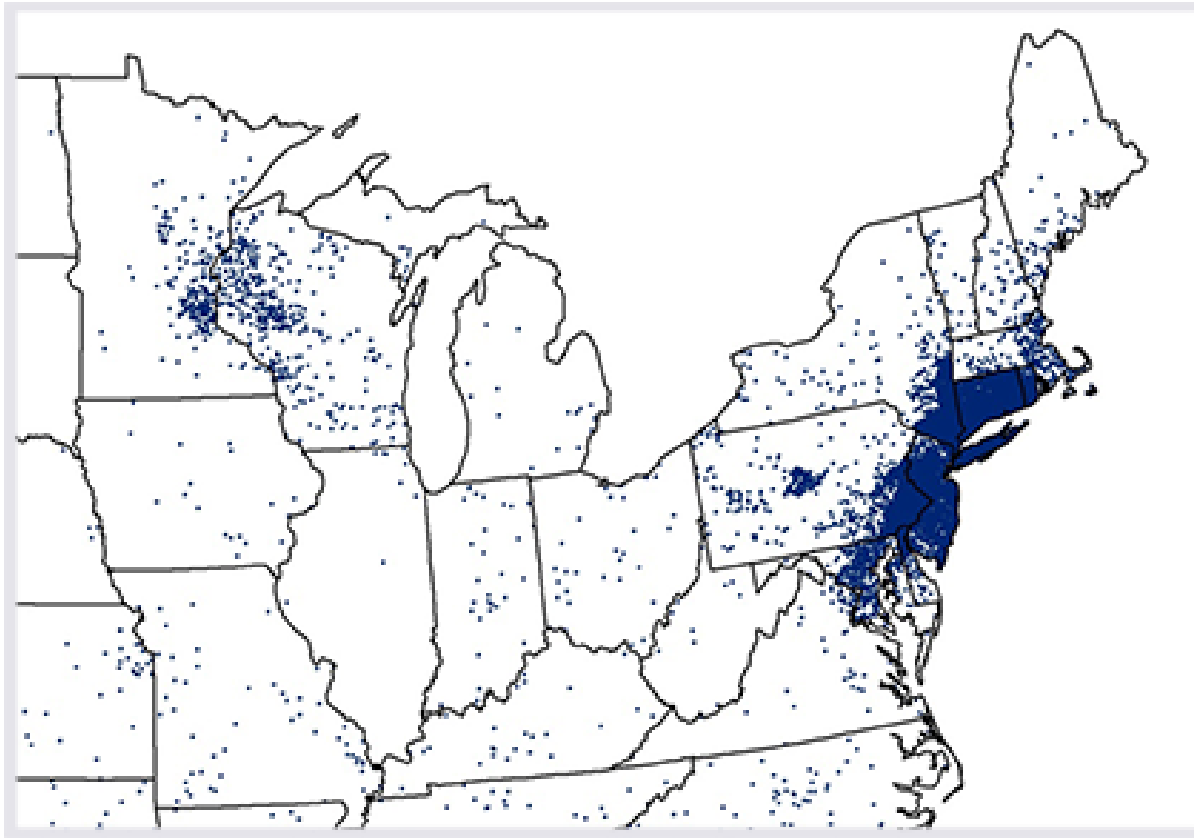
- Tobler's first law of Geography
 - *Everything is related to everything else, but near things are more related than distant things*
 - Values at locations near each other tend to be similar, with similarity decreasing with distance
 - Implies that phenomena are not distributed randomly (throughout space)
 - Imagine how the world would appear if everything were randomly distributed!



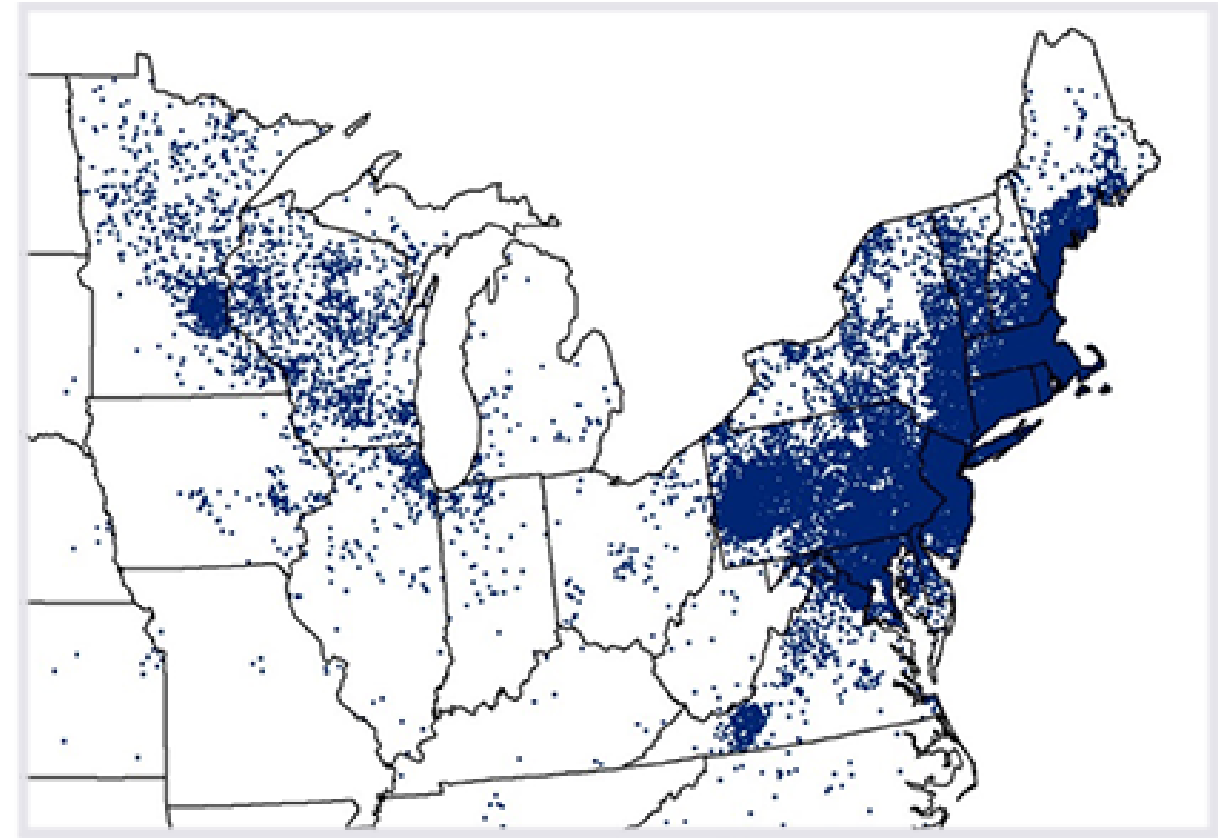
Geographic Variation

- Variation in some phenomenon across space or from place to place
 - We can observe this in tables, but view it in maps
 - Events (e.g., disease cases)
 - Locations (e.g., hospitals)
 - Values (e.g., average income)

Events



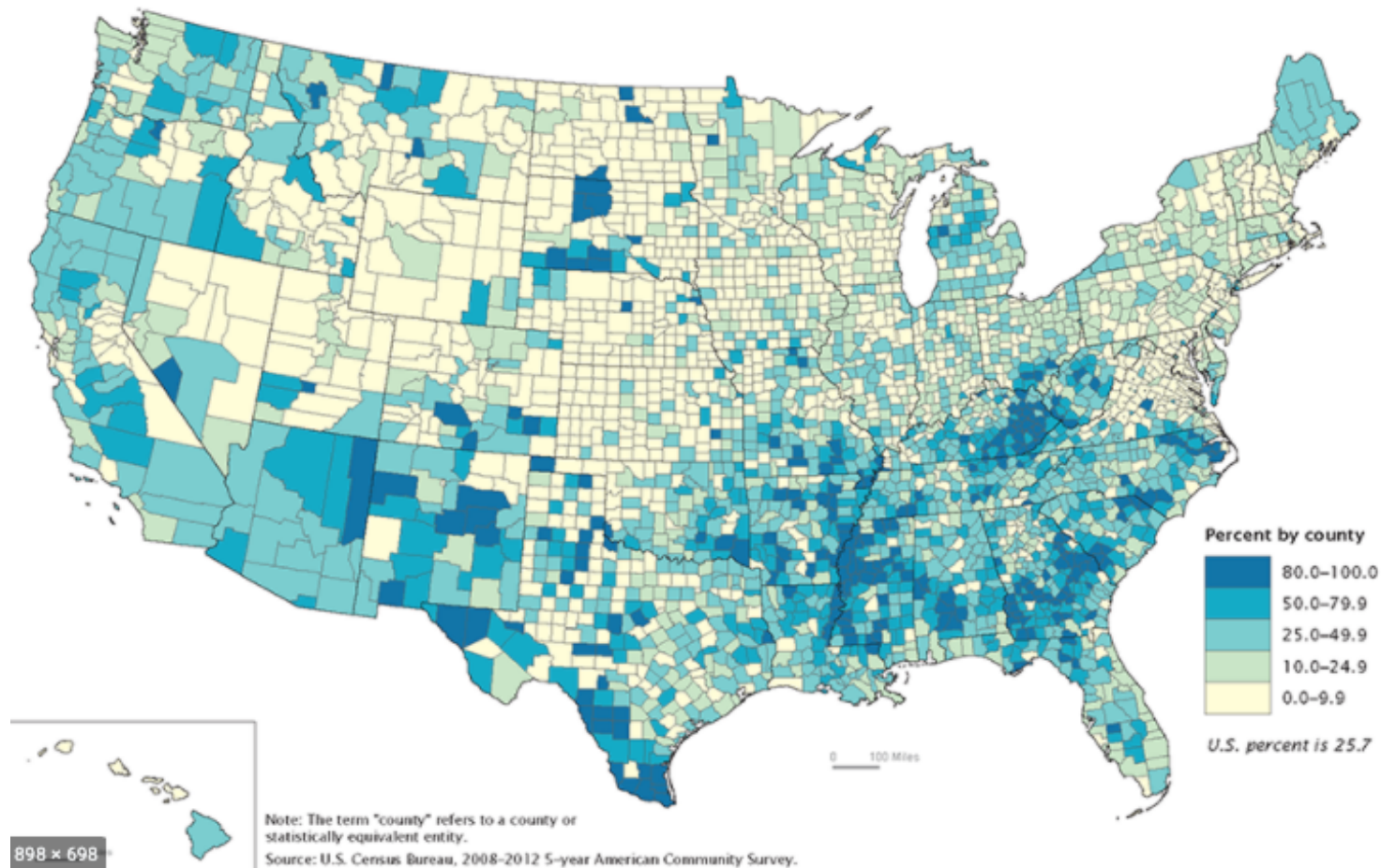
1996



2014

Lyme Disease Cases

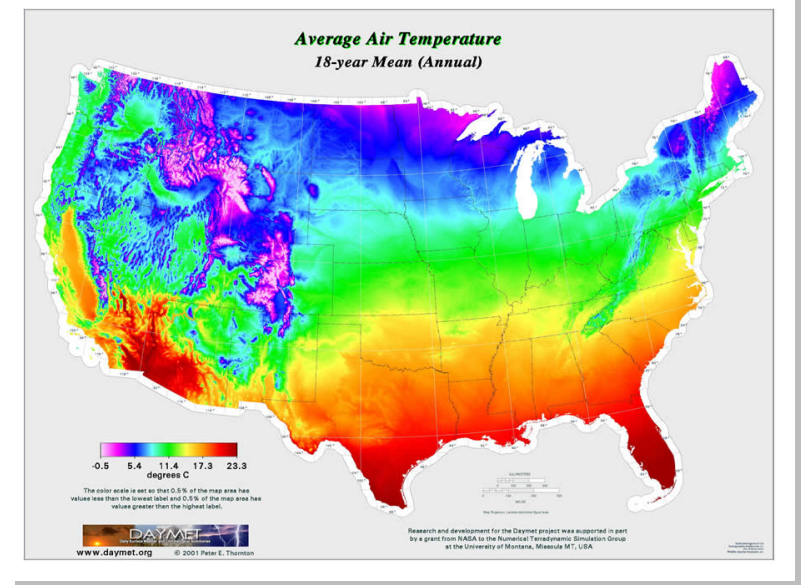
Values



% living under poverty

Geographic Variation

- Visual observation is subjective
 - While patterns may be visible, our eyes do not provide an objective test
 - We have techniques that assist us in interpreting spatial patterns



Geographic Variation

- Binary vs continuous
 - We can have various degrees of clustering or dispersion
 - We have techniques that assist us in quantifying degree of clustering



Spatial Cluster Analysis

- Two main approaches
 - Identifying whether events/values are clustered in space
 - Global, does not tell us “where”
 - Identifying clusters of events/values in space (deviations from expected)
 - Local regions having...
 - High/Low values
 - Higher density (unmarked points)

Point Data

- Two general types of point data
 - Unmarked points
 - A point is simply the location of an event
 - Every point is a "1"
 - Marked points
 - A point is the location of an event that has a measurable magnitude
 - The number of events at that particular location or the magnitude of the event at a particular location

Spatial Pattern

- Definition of Pattern (from Merriam-Webster)
 - The regular and repeated way in which something happens or is done
 - A natural or chance configuration
- Spatial Pattern
 - Arrangement of objects in space
 - One of the main concerns for geographers and geographical analysis
 - Understanding the distribution of phenomena throughout a landscape (people, objects, features)... and why this is/isn't important

Spatial Pattern

Events

- Basic Concepts

- Clustered

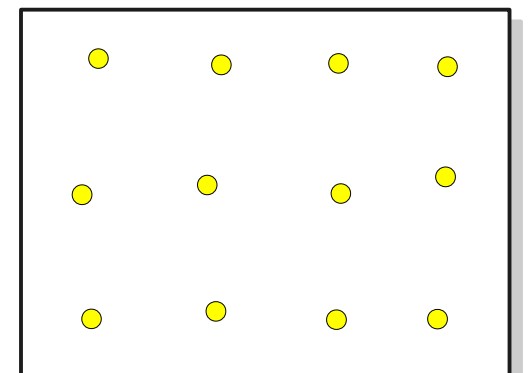
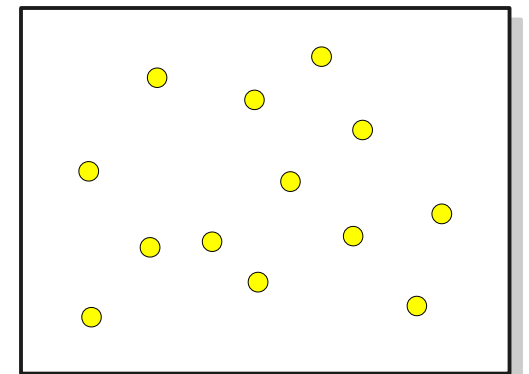
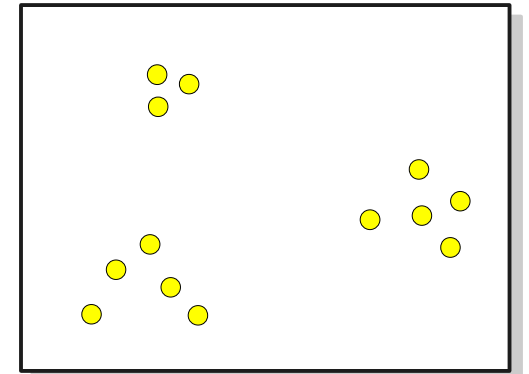
- Events are located or distributed near to one another

- Random

- Events are located or distributed such that there is no regular pattern

- Ordered (dispersed)

- Events are located or distributed in a regular or repeating fashion



Spatial Pattern

Attribute Values

- Basic Concepts

- Clustered

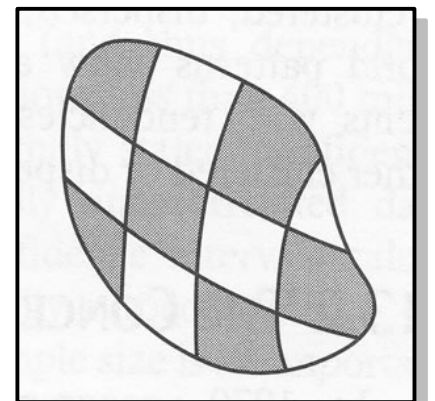
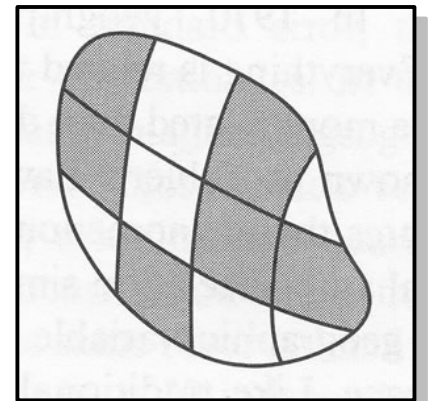
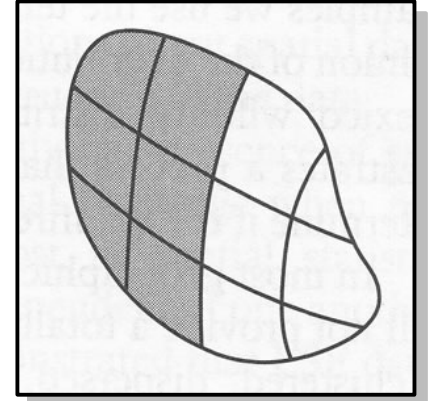
- Values are configured or distributed near to one another

- Random

- Values are configured or distributed such that there is no regular pattern

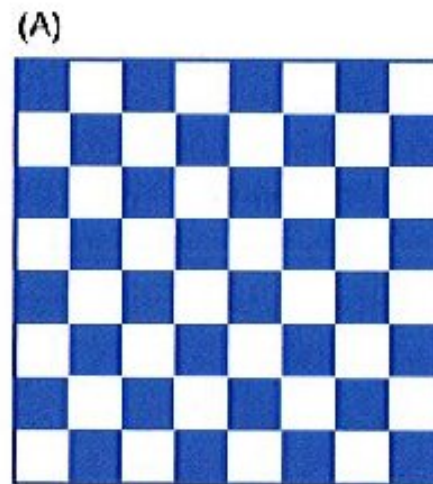
- Ordered (dispersed)

- Values are configured or distributed in a regular or repeating fashion

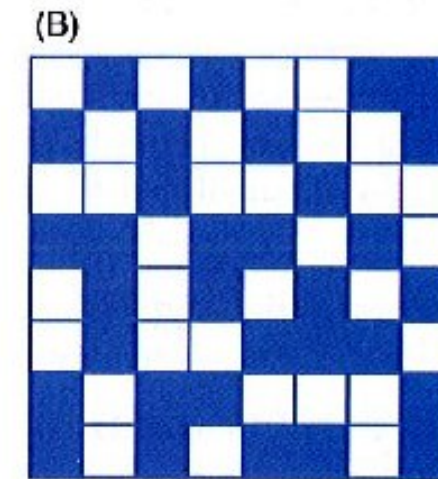


Polleverywhere

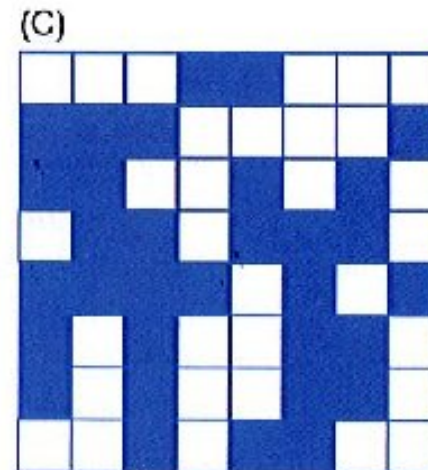
- How good are you at perceiving patterns?
- pollev.com/goelvarun553



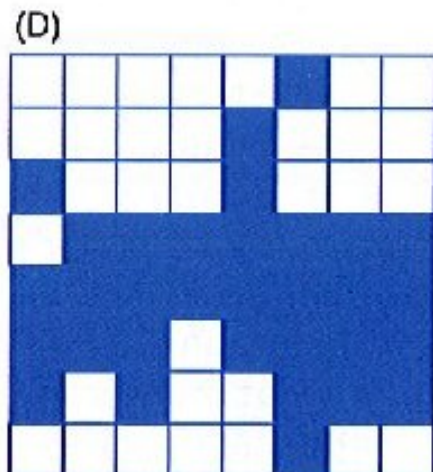
$I = -1.000$
 $n_{BW} = 112$
 $n_{BB} = 0$
 $n_{WW} = 0$



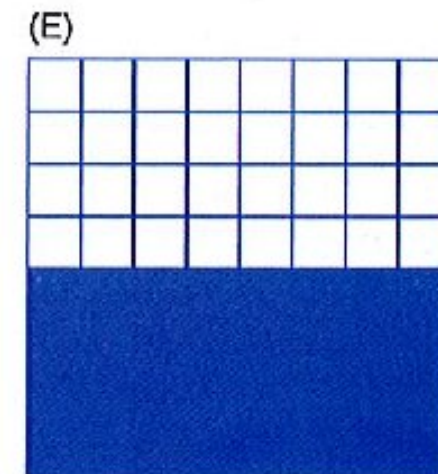
$I = -0.393$
 $n_{BW} = 78$
 $n_{BB} = 16$
 $n_{WW} = 18$



$I = 0.000$
 $n_{BW} = 56$
 $n_{BB} = 30$
 $n_{WW} = 26$



$I = +0.393$
 $n_{BW} = 34$
 $n_{BB} = 42$
 $n_{WW} = 36$



$I = +0.857$
 $n_{BW} = 8$
 $n_{BB} = 52$
 $n_{WW} = 52$

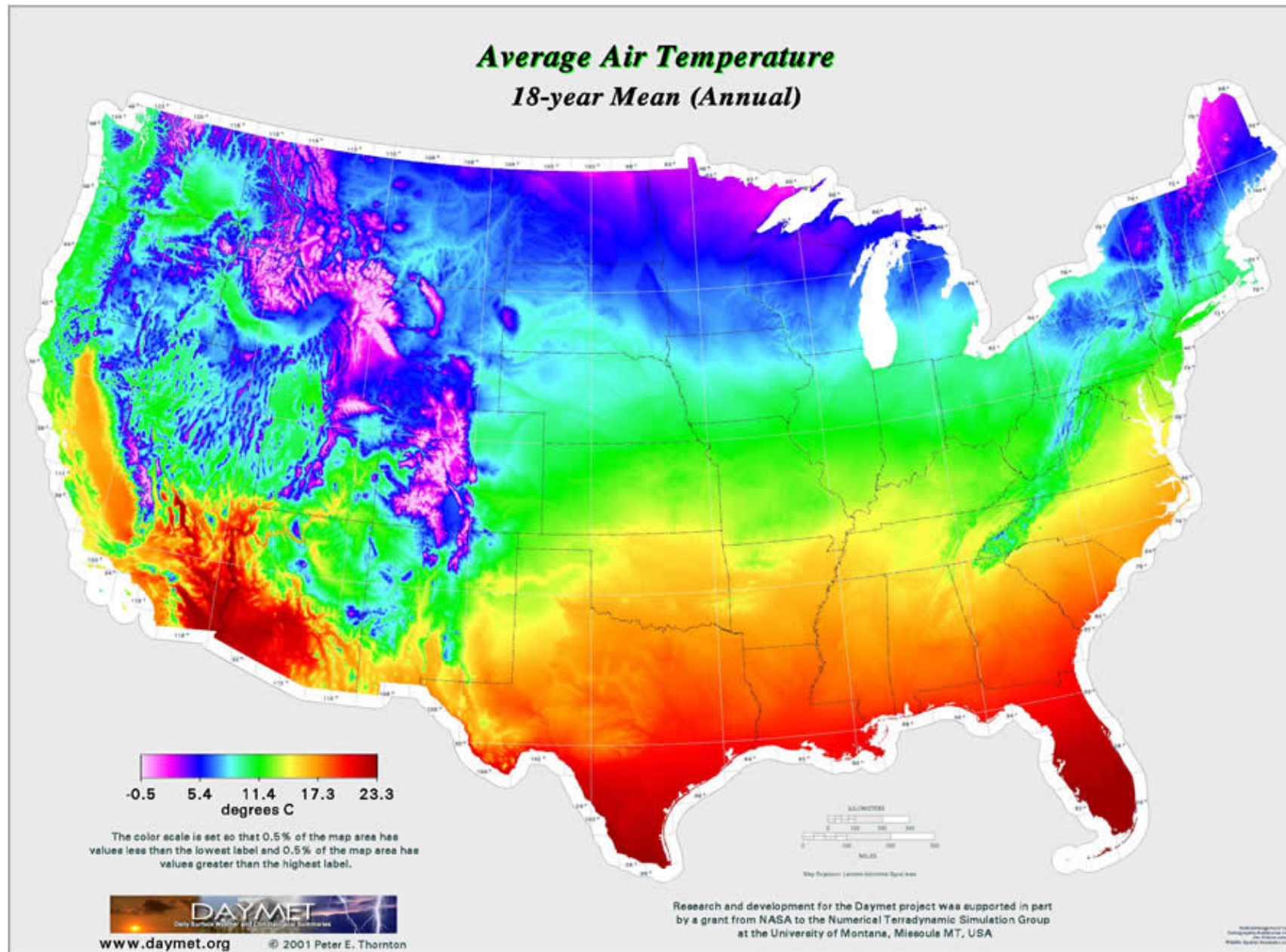
Figure 4.1 Field arrangements of blue and white cells exhibiting: (A) extreme negative spatial autocorrelation; (B) a dispersed arrangement; (C) spatial independence; (D) spatial clustering; and (E) extreme positive spatial autocorrelation. The values of the I statistic are calculated using the equation in Section 4.6 (*Source: Goodchild 1986 CATMOG, GeoBooks, Norwich*)

Spatial Autocorrelation

- Spatial Autocorrelation
 - The degree of similarity between objects that are located near each other
 - The arrangement or pattern of “values” within the landscape
 - Clustered, Random, Dispersed
 - Can be measured, quantitatively
 - For an entire region (global)
 - In a smaller area within the region (local)

Spatial Autocorrelation

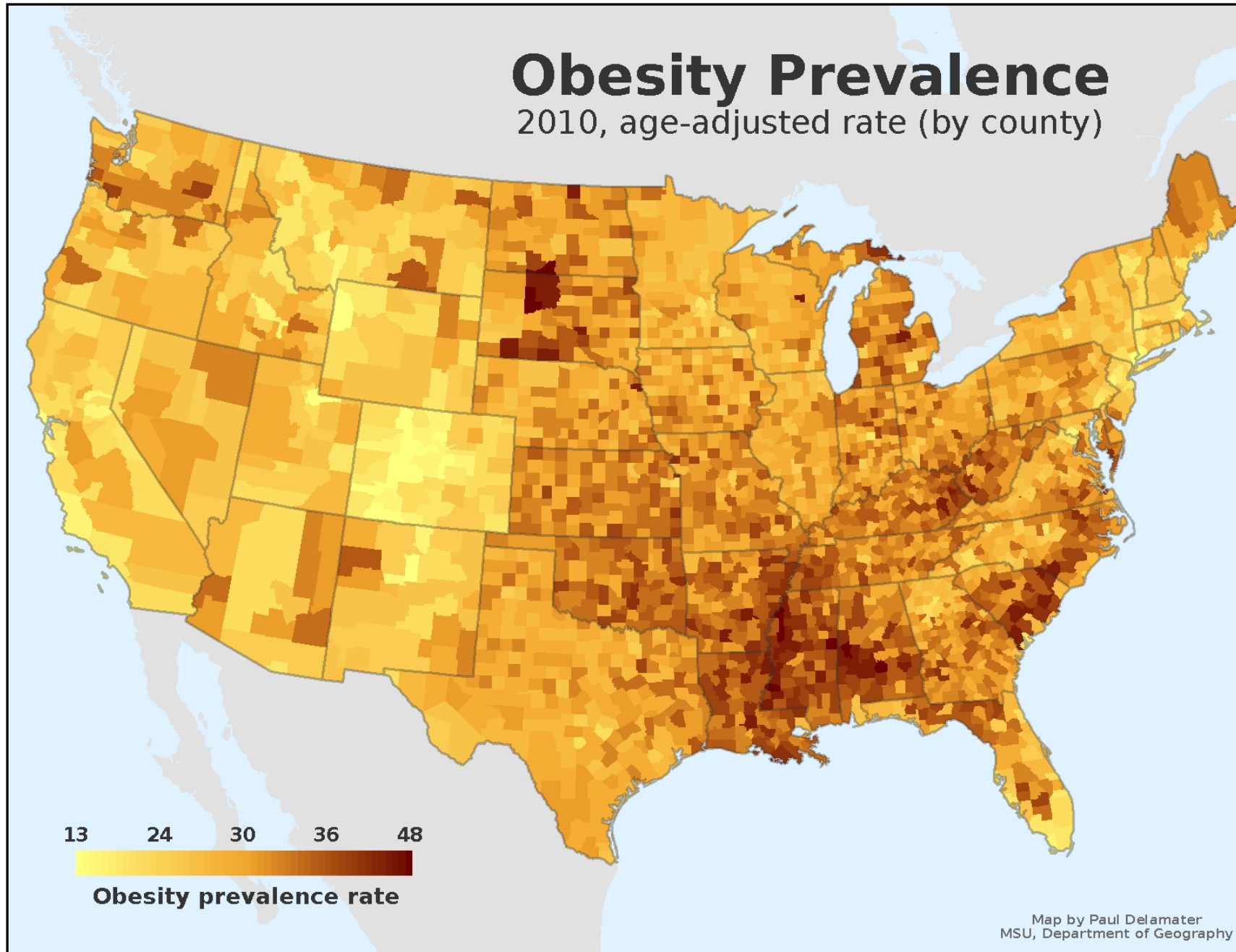
- For areal (polygon), point, or raster data, we measure how variable values are arranged
 - Not simply the locations of the objects, but the attributes associated with them
 - Not recommended for count data unless population generating counts are exactly the same from place to place



http://www.crwr.utexas.edu/gis/gishydro05/Time/daymet_files/ta_a_pa_letter.jpg

Obesity Prevalence

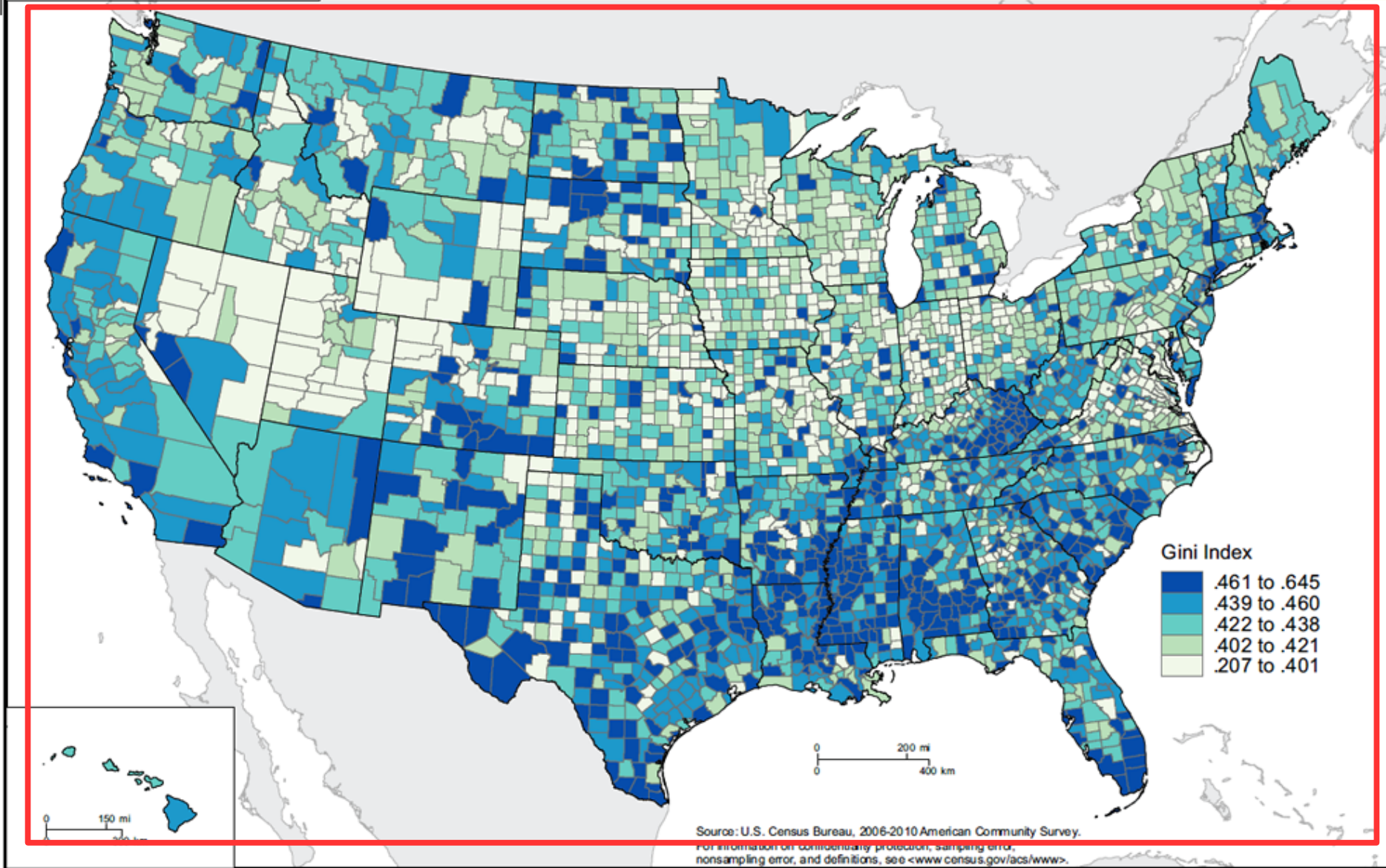
2010, age-adjusted rate (by county)



INCOME INEQUALITY

Figure 1.
Quintiles of Gini Index by County: 2006–2010

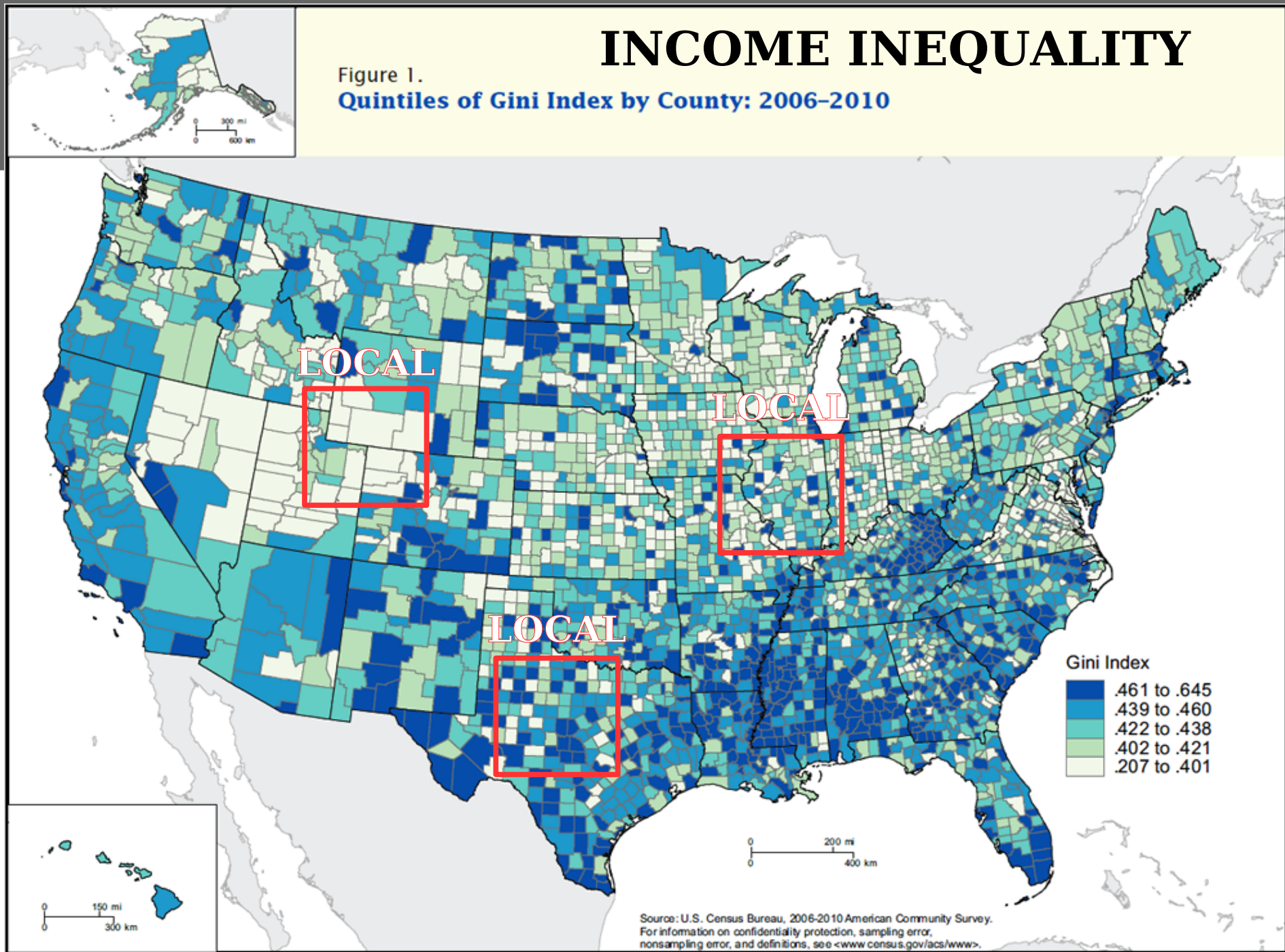
GLOBAL



<http://irjci.blogspot.com/2012/03/how-unequal-is-household-income-in-your.html>

INCOME INEQUALITY

Figure 1.
Quintiles of Gini Index by County: 2006–2010



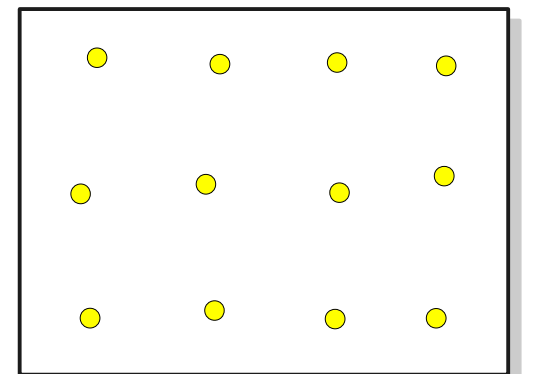
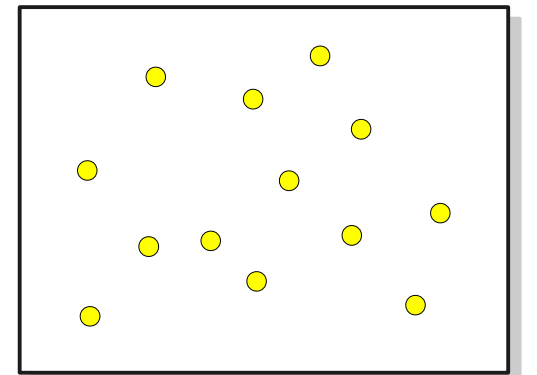
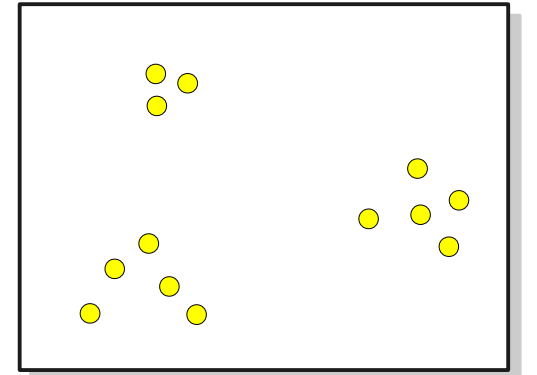
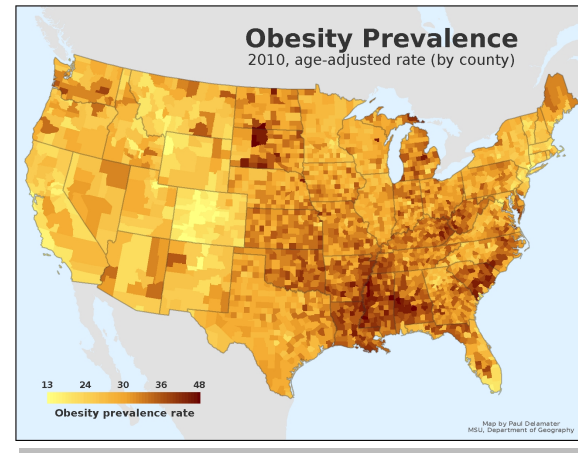
<http://irjci.blogspot.com/2012/03/how-unequal-is-household-income-in-your.html>

Spatial Autocorrelation

- Spatial Autocorrelation
 - The degree of **similarity** between objects that are located **near** each other
 - **Requires the definition of neighbors to evaluate similarity**

Spatial Autocorrelation

- Spatial Autocorrelation
 - Note that this is **self** similarity
 - Only a single variable or set of events!
 - Correlation *between variables* is something different



Correlation

- Correlation
 - The relationship between things that happen or change together
 - Other terms used: relationship, association, interaction
 - Using statistical techniques, we can measure and test the relationship between things

Pattern and Process

- A spatial pattern is generally the result of some spatial process
 - Clustering (**auto**correlation) approaches tell us about the *patterns* we see
 - We map and describe patterns
 - We use statistical models to understand or explain *processes*
 - One example is **bivariate correlation**

Keywords

- Geographic variation
- Spatial cluster analysis
- Unmarked and marked points
- Spatial pattern
- Clustered, random, dispersed
- Spatial autocorrelation
 - Global, local
- Correlation