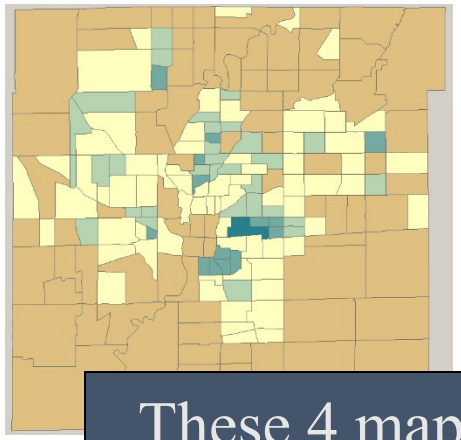
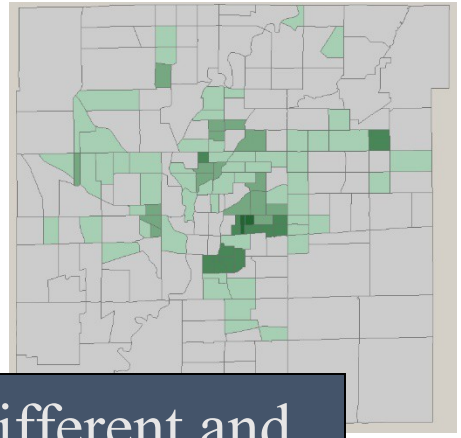
Data Classification Methods
What's the Difference?

Standard Deviation

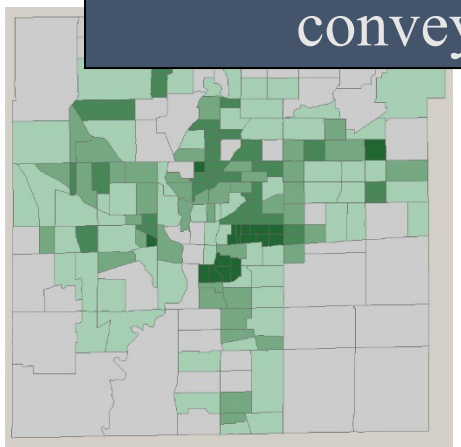


Equal Interval

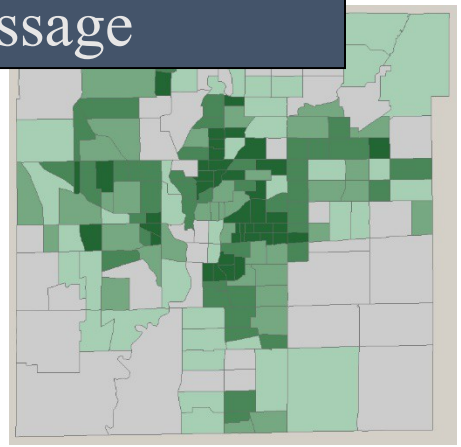


These 4 maps are all quite different and convey a different message

Natural Breaks



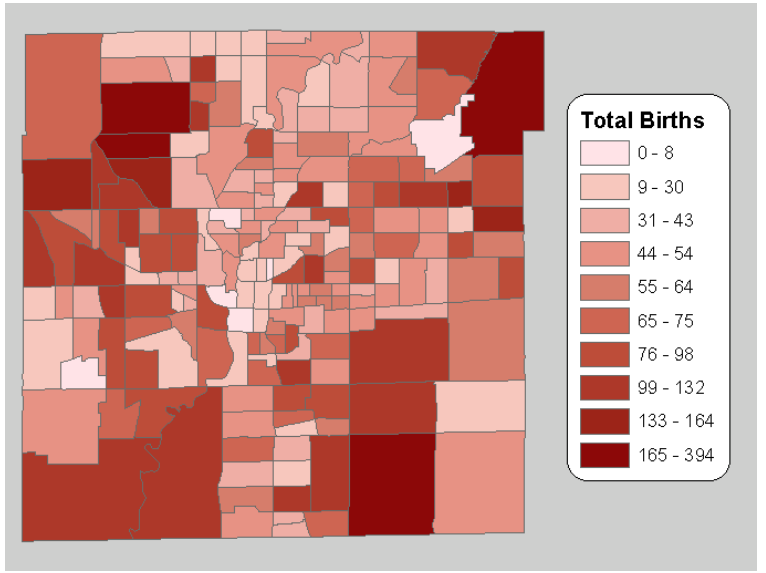
Quantiles

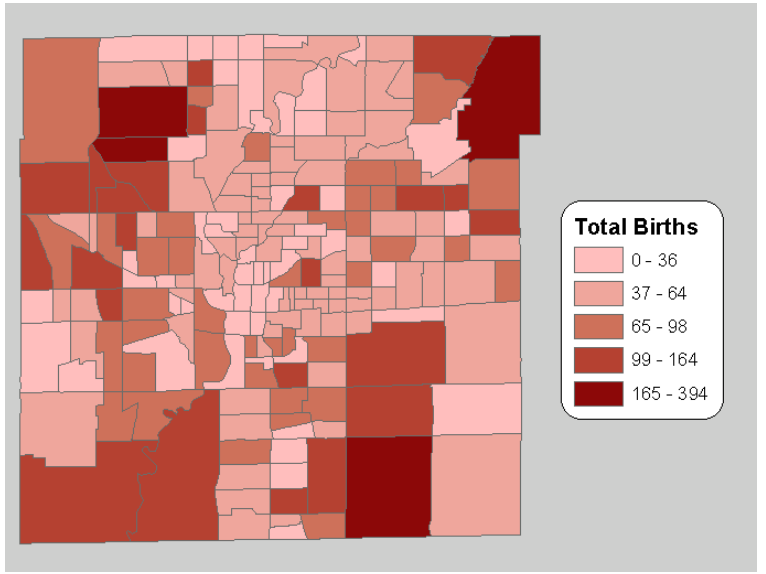


Data Classification

- Values are grouped into classes in order to simplify mapped patterns for the reader.
- No more than 6 and no less than 4 classes are recommended.

> 3 classes < 7





Data Classification Methods
Equal Interval

Encloses equal amounts of the range of the mapped data within each class interval.

For example, if you need four classes for a dataset with values that fall within a range of 0 to 100, the classes would be:

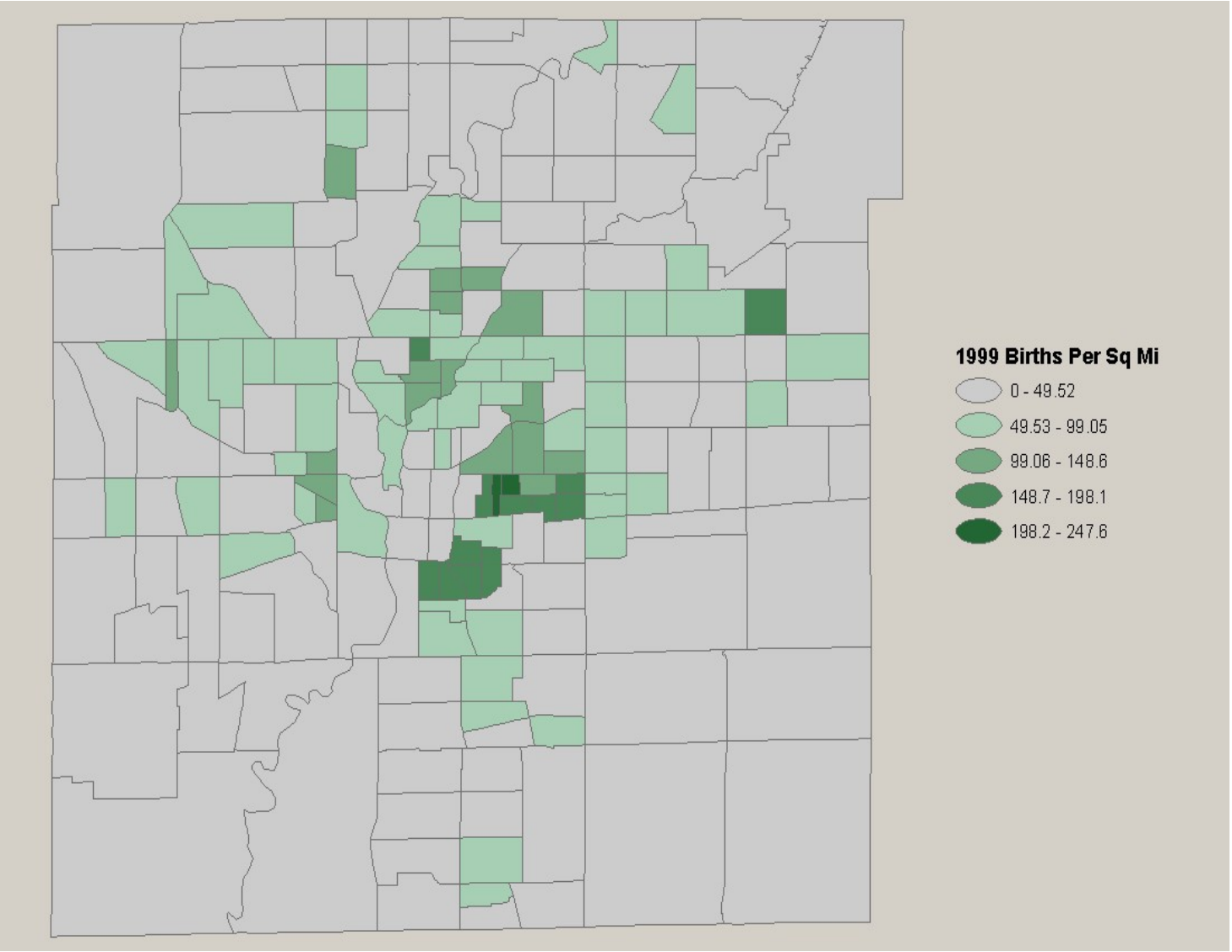
0 to 24

25 to 49

50 to 74

75 to 100

Equal Interval Example



Data Classification Methods
Equal Interval Classes – How it Works

- GIS subtracts lowest value in data set from the highest.
- It divides the number by the number of classes you specified.
- It adds the number to the lowest value to get the maximum for the first class.
- It then adds the maximum value to set the breaks for the rest of the classes.

Data Classification Methods
Equal Interval Classes – What Are They Good For?

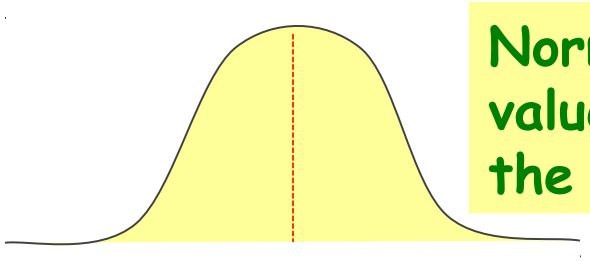
- Presenting information to a non-technical audience since equal intervals are easier to interpret since the range of each class is equal.
- Mapping continuous data – such as precipitation or temperature.

Data Classification Methods
Equal Interval Classes - Disadvantages

If data values are clustered rather than evenly distributed, there may be many features in one or two classes and some classes may have few or no features.

Data Classification Methods
Standard Deviation

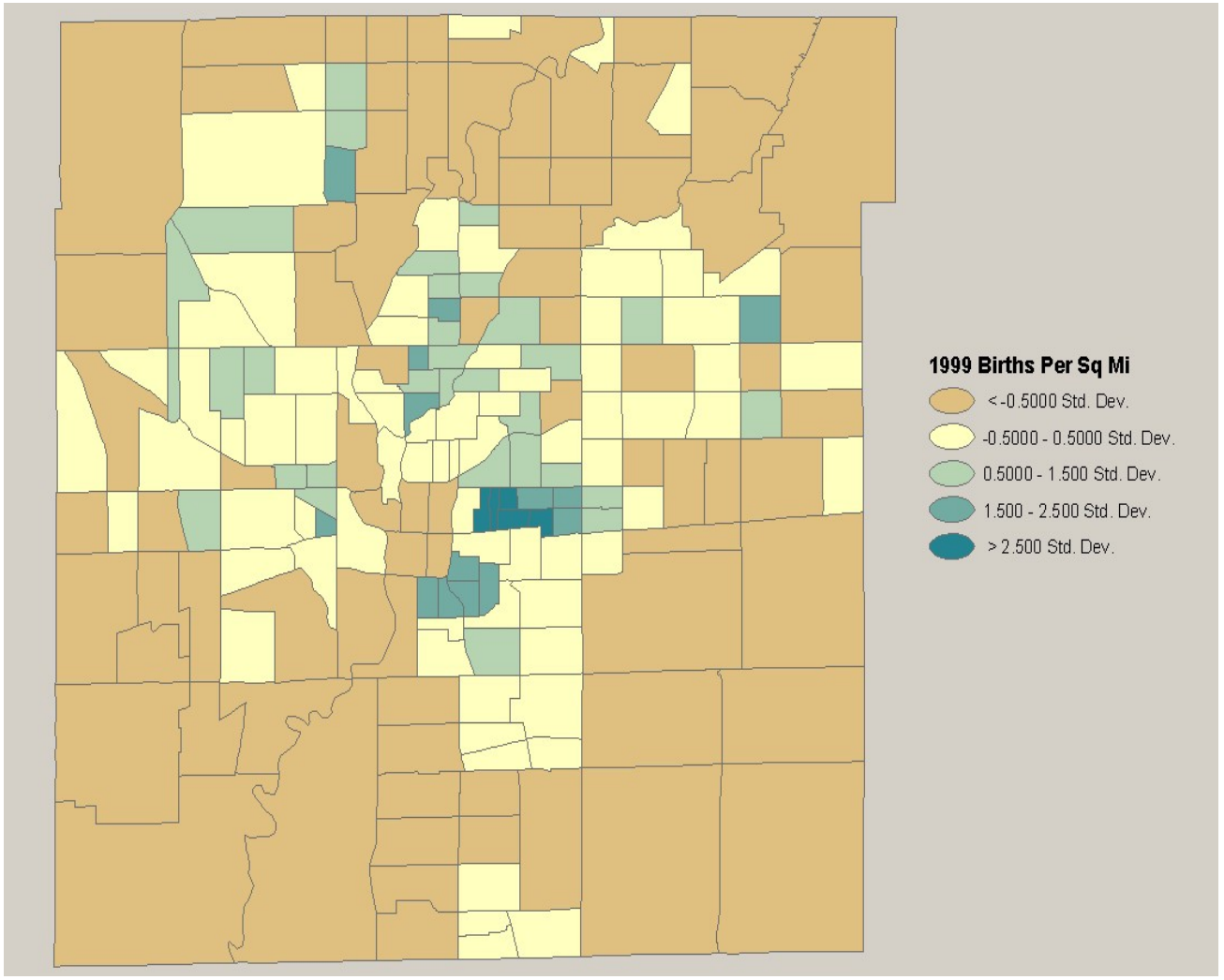
- This classification method should only be used when the data set approximates a normal distribution.



Normal distribution has values clustered around the mean

- Class intervals created with this method should only be used when the reader understands them

Standard Deviation Example



Data Classification Methods
Standard Deviation: How It Works

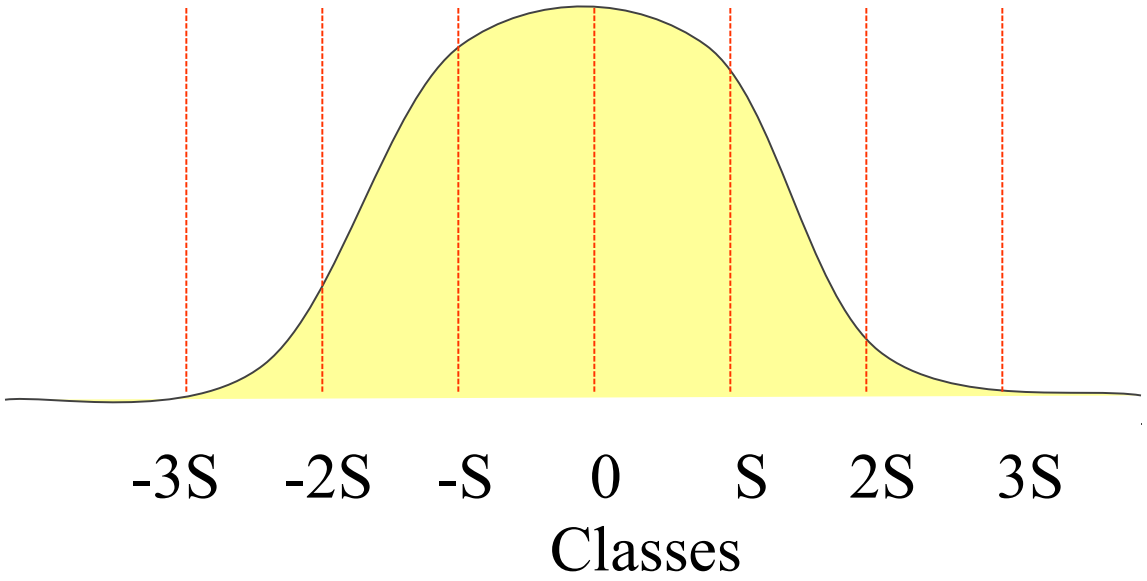
1. The GIS finds the mean value by adding all data values and dividing by the number of features.
2. It then calculates the standard deviation (s) by subtracting the mean from each value and squaring it (to make it positive), sums the numbers and divides by the number of features.
3. It then takes the square root to find the final standard deviation.
4. GIS then creates classes above and below the mean based on the number of standard deviations you specify such as ½ or 1 standard deviations.

Data Classification Methods:
Standard Deviation: How It Works

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Data Classification Methods:
Standard Deviation: How It Works

The data classes are centered on the mean



Data Classification Methods
Standard Deviation: What is it good for?

- Seeing what features are above or below an average value.
- Displaying data that has many values around the mean and few further from the mean (this is statistically called a bell curve or normal distribution).

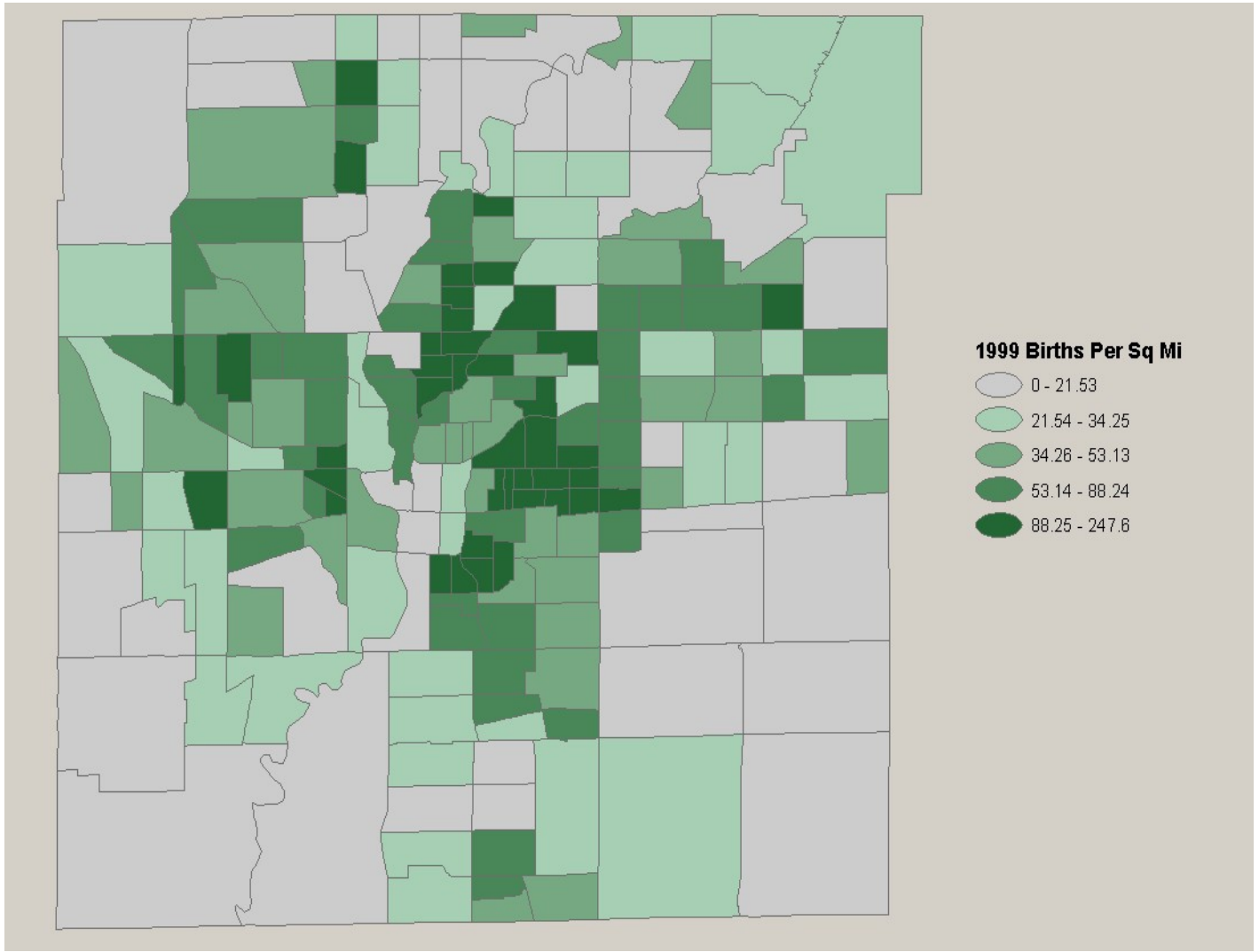
Data Classification Methods
Standard Deviation: Disadvantages

- Map does not show actual values of features, only how far they are from the mean.
- Very high or low values (called outliers) can skew the mean so that most features fall in the same class.

Data Classification Methods
Quantiles

- Quantiles produces irregular intervals.
- Developing class boundaries with quantiles assures an equal number of values in each class and minimizes the importance of the class boundaries.

Quantiles Example



Data Classification Methods
Quantiles: How It Works

1. The GIS orders the features, based on attribute value, from top to bottom and sums the number of features as it goes.
2. It divides the total number of features by the number of classes you specify to get the number of features in each class.
3. It assigns the first features in the order to the lowest class until that class is filled and then moves on to the next lowest class, and so forth.

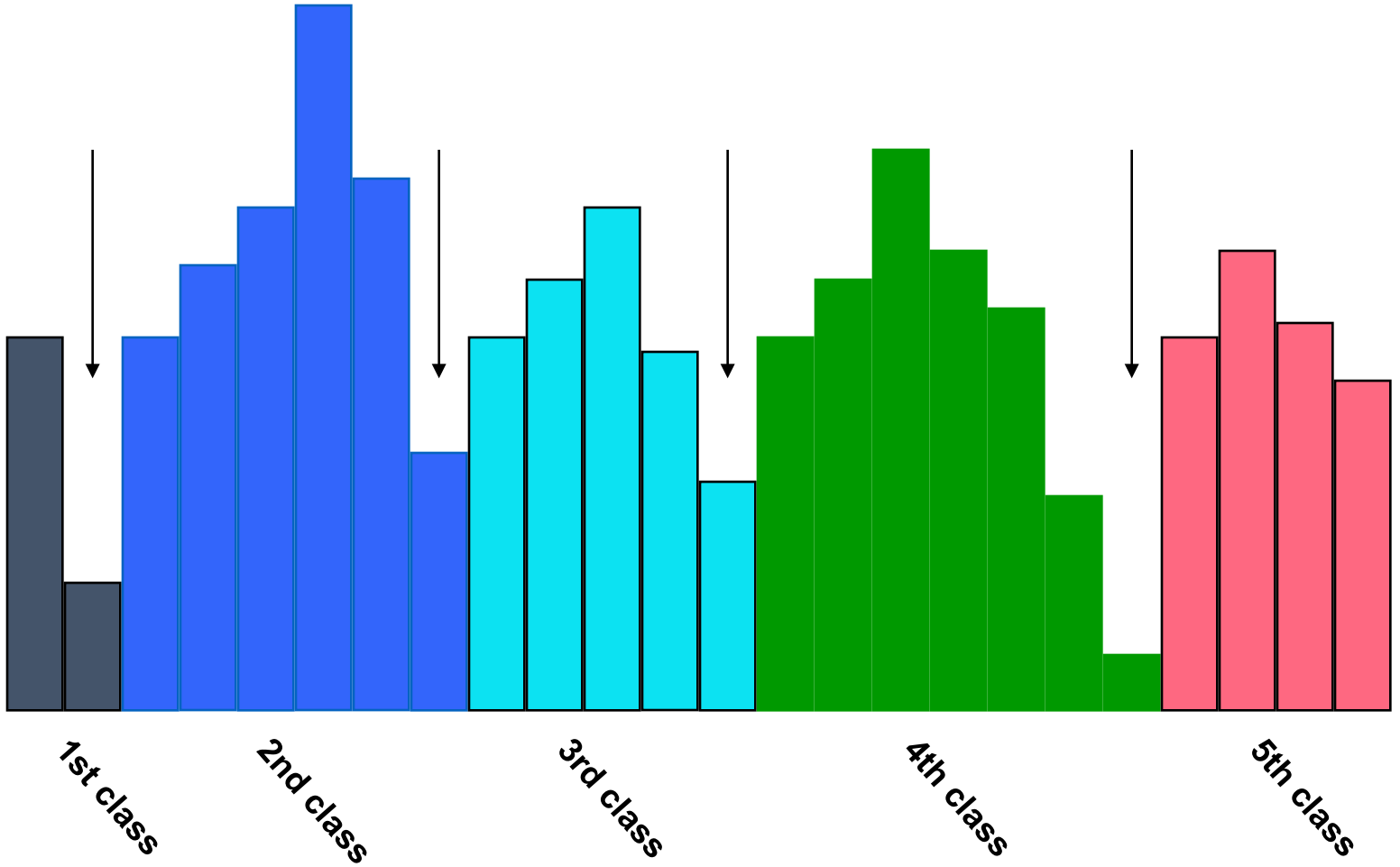
Data Classification Methods
Quantiles: Disadvantages

- Features with close values may end up in different classes, especially if values cluster. This may do one of two things:
  - exaggerate the differences between features
  - A few widely ranging features may end up in the same class, thus minimizing the differences between features
- If areas vary greatly in size, a quantile distribution may skew the patterns on the map.

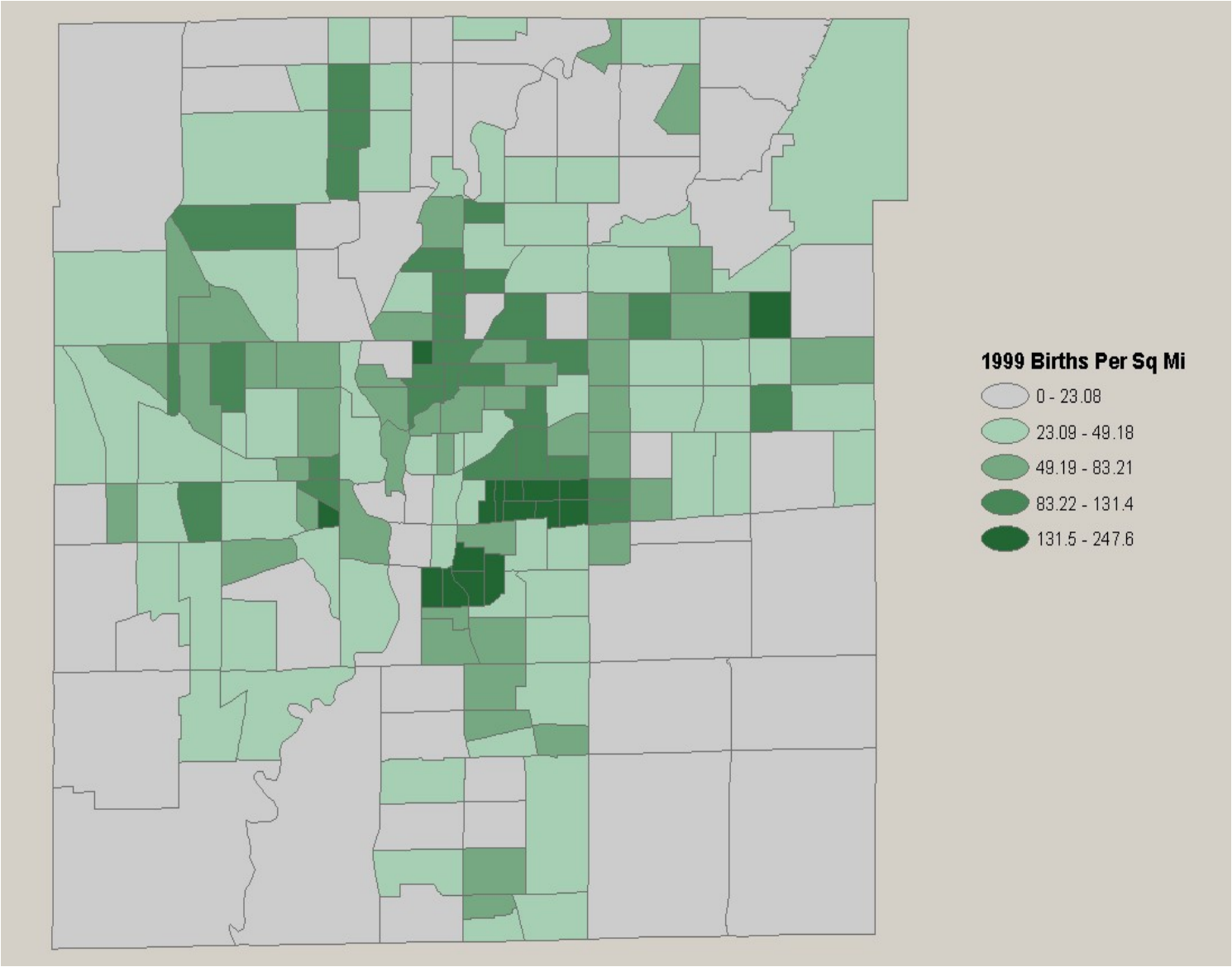
Data Classification Methods
Natural Breaks

- Natural breaks are determined with a frequency histogram. Class boundaries are identified as troughs in the data.
- Selection of class boundaries tends to place large numbers of similar values in the same class.
- Many data sets will not have obvious natural breaks which means that a map created with this method would tend to show breaks where none really exist.

Data Classification Methods
Natural Breaks



Natural Breaks Example



Data Classification Methods
Natural Breaks: How it Works

1. The GIS finds groupings and patterns that are inherent in your data.
2. Data values that cluster are put in a single class.
3. Class breaks are defined where there is a gap between clusters of values.

Data Classification Methods
Natural Breaks: Disadvantages

- Since the class values are specific to individual classes, it is difficult to compare the map to other maps.
- Choosing the number of classes can be difficult especially if the data is evenly distributed.

What to Do About Outliers

- Put them in their own class
- Group them into a single class
- Group them with the next closest class
- Draw them using a special symbol