Introduction to the Mapping Sciences

Map Composition & Design IV:

Measurement & Class Intervaling Principles & Methods

ROWAN UNIVERSITY

1

Overview:

- Levels of measurement
- Data classing principles and methods

2

Levels of Measurement

3

Levels of Measurement

- Nominal
- Ordinal
- Interval
- Ratio

4

Measurement Scales

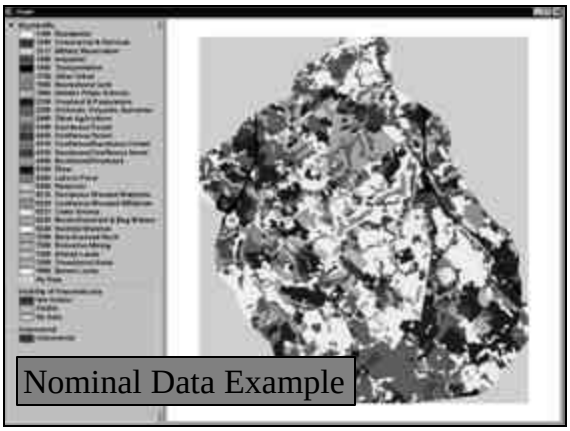
- **Nominal**: Categorical measure [e.g., land use map]
  - Lowest level [= only possible operation]
- **Ordinal**: Ranking Measure
  - Lowest quantitative level [=, >, <]
  - Strong ordinal: all objects placed in order, no ties.
  - Weaker ordinal: all objects placed in order, but ties exist.
  - Weakest ordinal: rating scales:
    - Strongly agree, agree, disagree, strongly disagree.

5

Nominal Categories

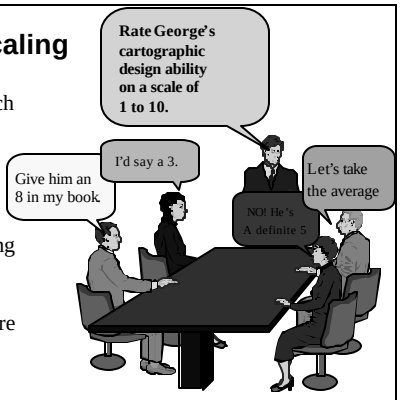
- Establish set of categories.
- Every object falls into a category.
- No object falls into more than one category.
- What rule to use for dividing space??

6



Ordinal Scaling

- **Strong:** rock hardness scratch test.
- **Weaker:** soil percolation categories.
- **Weakest:** rating resulting from independent scaling. Is there an underlying metric at all?



Interval Scale

- Can measure the size of a difference, but not how many times one observation is greater than another:
  - Temperature: Consider ratio of 80°F and 40°F
    - 80°F / 40°F = 2.00
    - 26.6°C / 4.44°C = 5.99

    } Temperature Scale Conversion from Degrees F to C
  - No true zero amount. At 0°F there is still some of the thing called temperature. But the interval of 20° ***is*** two times as large as the interval of 10°.
  - Operations: [=, >, <, +, -]

9

Ratio Scale

- Can place numbers in ratio to each other:
  - Edith makes three times as much money as Walter.
  - Bill lives twice as far from here as Mary.
    - 20 miles / 10 miles = 2.0
    - 32.3 km / 16.1 km = 2.0

    } Distance Scale Conversion from Miles to Kilometers
  - There is a true zero. You ***can*** have zero money. Ratio remains same through transformation of metric.
  - Operations: [=, >, <, +, -, *, /, ^]

10

Measurement Scale Pointers

- Scales are downwardly inclusive
  - OK to move down the scale, difficult or impossible to move up the scale.
- During analysis try to keep all measurement at the highest possible level.
- Be careful when performing operations that involve data layers of different scale levels.

11

Measurement Scale Summary

- **Nominal**
  - Categorical measure [e.g., land use map]
    - Lowest level: Operations [=]
- **Ordinal**
  - Ranking [strong or weak]
    - Lowest quantitative level [=, >, <]
- **Interval**
  - How big a difference, but not how many multiples [e.g., temperature data]
    - Operations [=, >, <, +, -]
- **Ratio**
  - Can put number in ratio [Mary lives twice as far as Bill]
    - Highest level: Operations [=, >, <, +, -, *, /, ^]

Data Classing

Principles & Methods

13

Overview of Map Classing

- What is classing? Why do we class?
- Overriding principles.
- Principles for deciding:
  - Number of classes.
  - Method to use in classing.
- Methods of Classing:
  - Natural Breaks
  - Equal Interval
  - Quantile
  - Mean and Standard Deviation
  - Arithmetic Progression
  - Geometric Progression

14

What Is Classing?

- Classification process to reduce a large number of ***individual*** quantitative values to:
  - A smaller number of ***ordered categories***, each of which encompasses a portion of the original ***data value range.***
- Various methods divide the ***data value range*** in different ways.
- Varying the method can have a very large impact on the look of the map.

15

Fundamental Principles

- **Each** of the original (actual) data values must fall into ***only of the classes***. Each data value has a class home.
- **None** of the original data values falls into ***more than one class***.
- These two rules are supreme: if any method results in classes that violate these rules, the resulting classes must be altered to conform to the two fundamental principles. ***Always***.

Always Don't you forget it

16

Shorthand Way of Saying This

- Classes must be:
  - Mutually exclusive

    AND
  - Exhaustive

17

Principles for Deciding on Number of Classes

- Rules of thumb:
  - Monochrome color schemes: no more than 5 to 7 classes.
  - Multi-hue map: no more than 9.
- BUT…it depends on other things such as ….

18

Principles for Deciding on Number of Classes

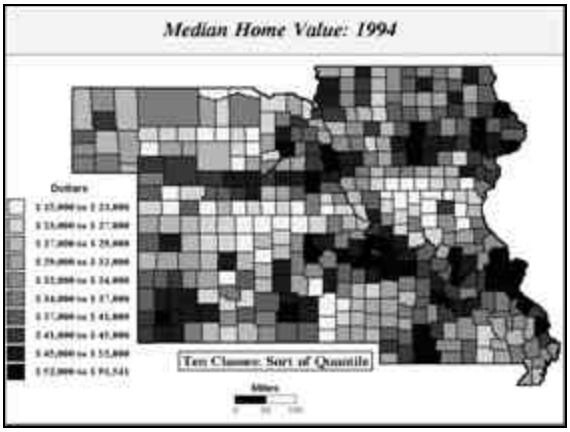
- Communication goal
- Complexity of Spatial Pattern
- Available Symbol Types

19

Principles for Deciding on Number of Classes

- Communication goal: **quantitative precision**: use larger number of class intervals. Each class will encompass a relatively small range of the original data values and will therefore represent those values more precisely.
- **Trade offs**:
  - Too much data to enable the ***information*** to show through.
  - Indistinct symbols.

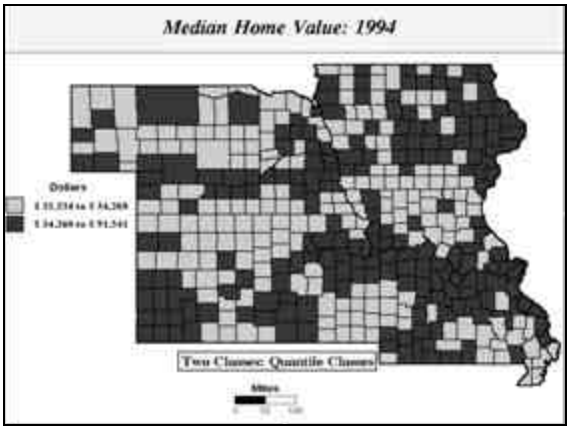
20



Principles for Deciding on Number of Classes

- Communication goal: **immediate graphic impact**: use smaller number of class intervals. Each class will be graphically clear, but will be imprecise quantitatively.
- **Trade offs**:
  - Potential for oversimplification
  - One class may include wildly varying data values

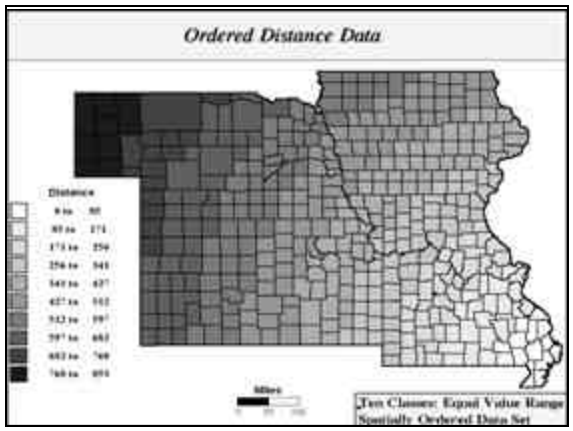
22



Principles for Deciding on Number of Classes

- **Complexity of Spatial Pattern**
  - Highly ordered spatial distribution can have more classes.
  - Complex pattern of highly interspersed data values requires fewer classes.

24



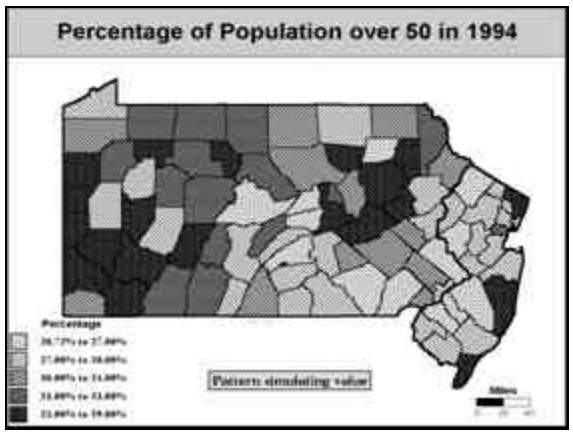
Principles for Deciding on Number of Classes

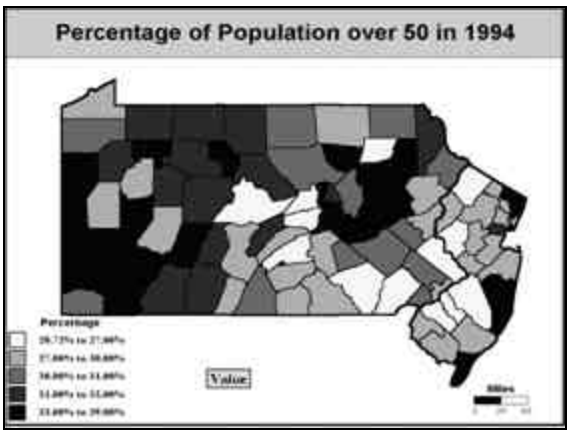
- Available Symbol Types:
  - Pattern
  - Value
  - Value + Pattern
  - Value + Pattern + Hue
  - Value + Pattern + Hue + Size

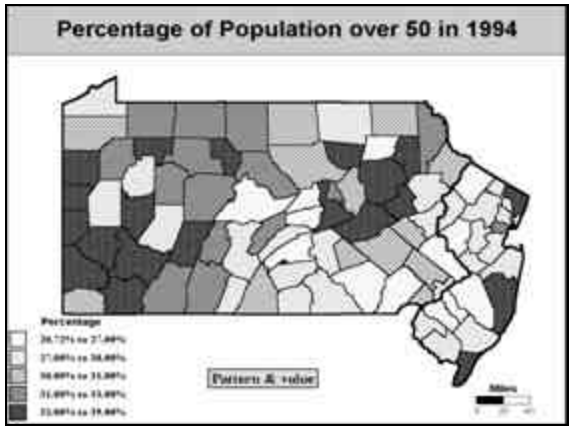
Fewer Classes

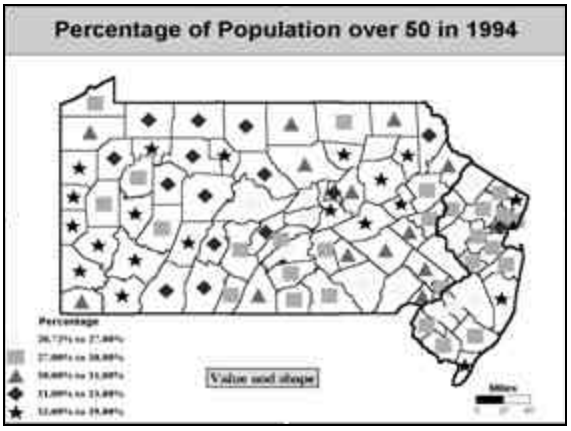
More Classes

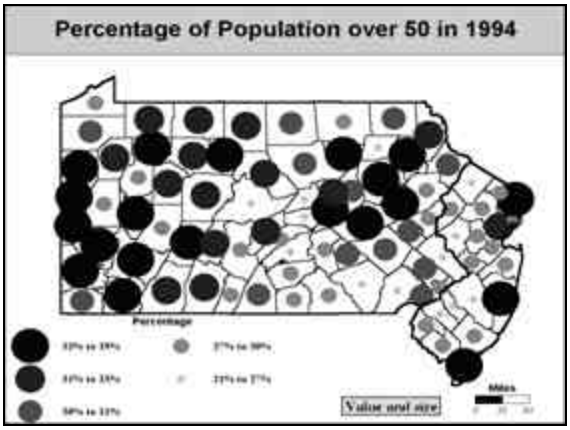
26

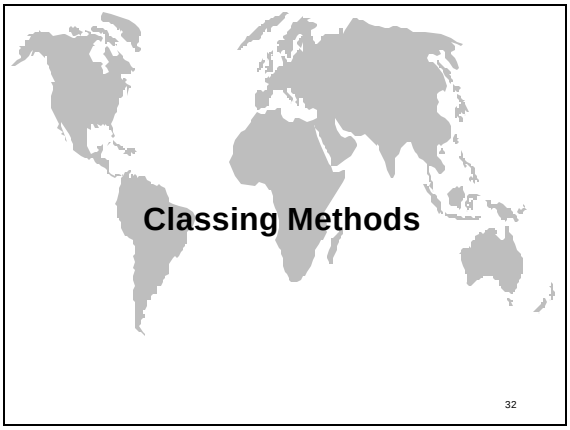










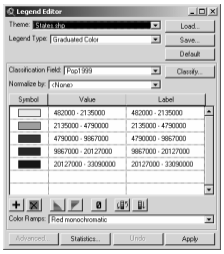


Natural Breaks

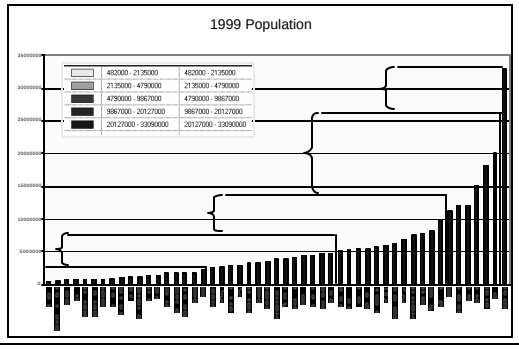
- Uses iterative analysis of variance.
- Attempts to create class breaks such that there is ***minimum*** variation in value within classes and ***maximum*** variation in value between classes.
- Arc View's default method

33

Natural Breaks



Natural Breaks



Natural Breaks

- Advantages
  - Maximizes the similarity of values within each class.
  - Increases the precision of the map given the number of classes.
- Disadvantages
  - Class breaks often look arbitrary.
  - Need to explain the method.
  - Method will be difficult to grasp for those lacking a background in statistical methods.

36

Equal Interval

- Each class encompasses an equal portion of the original data range. Also called equal size or equal width.
- Calculation:
  - Determine ***range*** of original data values:
    - Range = Maximum - Minimum
  - Decide on ***number of classes,*** **N.**
  - Calculate ***class width:***
    - **CW = Range / N**

37

Equal Interval

Class	Lower Limit	Upper Limit
1	Min	Min + CW
2	Min + CW	Min + 2CW
3	Min + 2CW	Min + 3CW
. . .	. . .	. . .
N	Min + (N-1) CW	Max

Equal Interval

- Example: Min = 0; Max = 100; N = 5
- Range = 100 - 0 or 100
- Class Width = 100 / 5 or 20

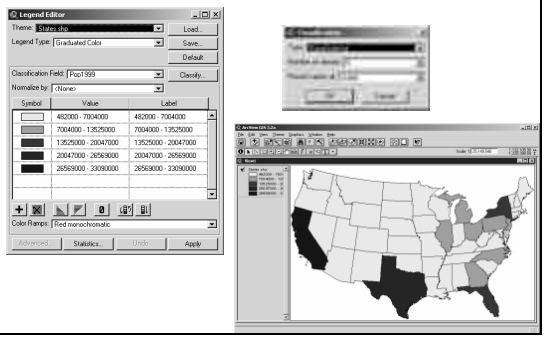
Class	Lower Limit		Upper Limit
1	**0**	**to**	**20**
2	**20**	**to**	**40**
3	**40**	**to**	**60**
4	**60**	**to**	**80**
5	**80**	**to**	**100**

Equal Interval

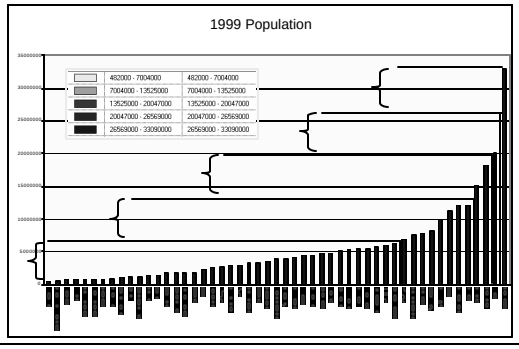
- Advantages:
  - Easy to understand, intuitive appeal
  - Each class represents an equal range or amount of the original data range.
  - Good for rectangular data distributions
- Disadvantages:
  - Not good for skewed data distributions-nearly all values appear in one class.

40

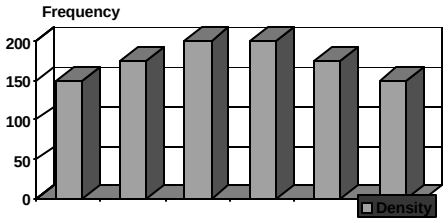
Equal Interval

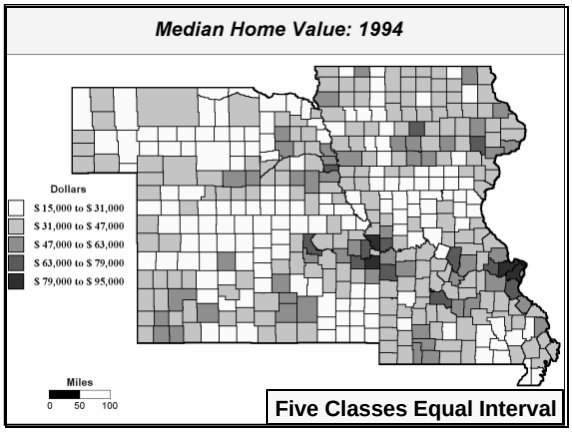


Equal Interval

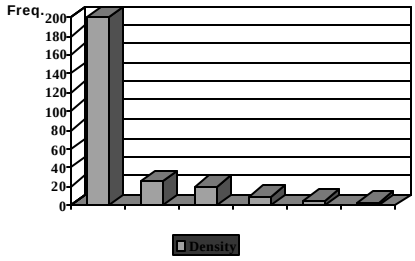


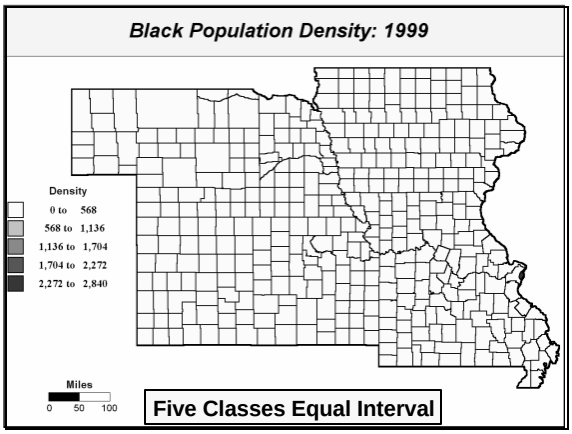
More Rectangular Data Distribution

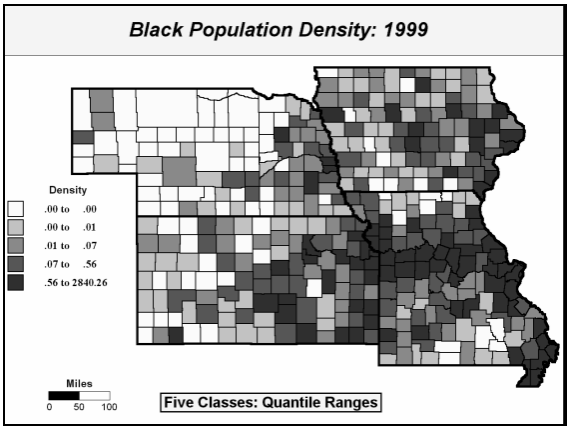




Skewed Data Distribution







Quantile Classes

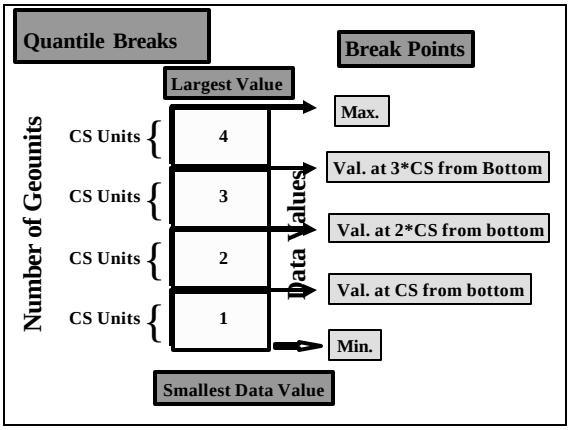
- Places an ***equal number of cases*** in each class.
- Sets class break points wherever they need to be in order to accomplish this.
- May not always be possible to get exact quantiles:
  - Number of geounits may not be equally divisible by number of classes. [21 / 5].
  - Putting same number of cases in each class might violate mutually exclusive classes rule.
    - 12 values in 4 classes:
      0 0 0 |0 0 3| 4 4 5| 6 7 7 **NO**
    - 0 0 0 0 0 |3| 4 4 5| 6 7 7 **YES**

48

Quantile Classes

- Decide on number of classes, **N**
- Determine the size of each class **[CS]** in number of geounits:
  - **Class Size = Total number of geounits / N**
- **Rank** geounits by data value: largest at top of list, next largest, and so on until you have the smallest at bottom of list.
- Count up from the bottom of list until you are at the geounit that is CS units from the bottom.
- ***Data Value*** at this point is the first class break point.
- Count up CS more units to find second break point.
- Continue until you have N-1 break points.
- Remember: mutually exclusive & exhaustive classes.

49

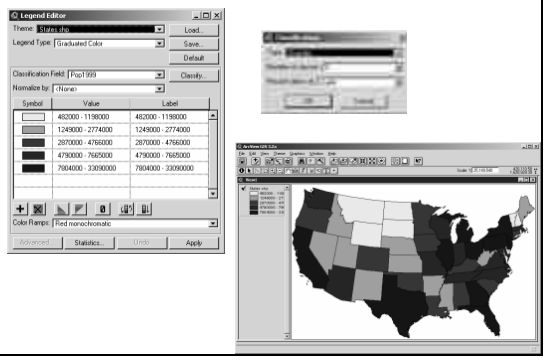


Quantile Example

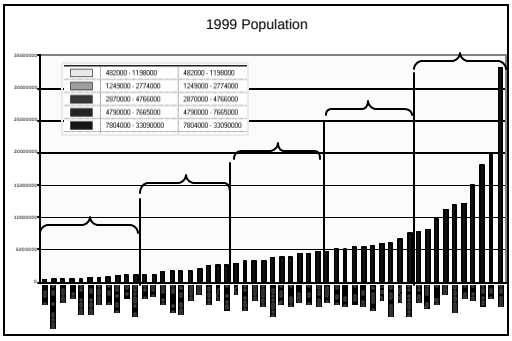
- **Data Values**
  10, 7, 9, 15, 8, 6, 4, 8, 10, 2, 8, 3
- **Number of Classes = 3**
- **Number of geounits = 12**
- **Class Size = 12 / 3 or 4**
- **Ranked Data Values**

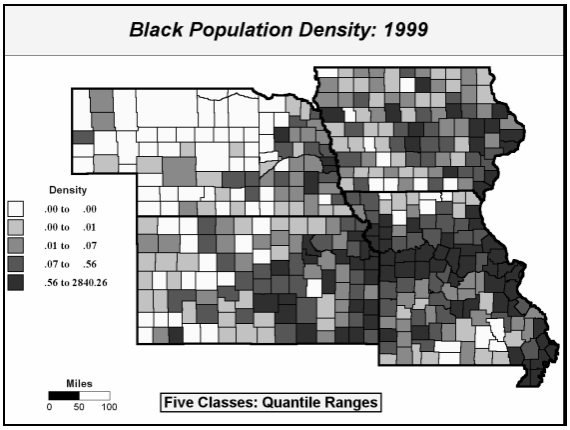
15 → **Max.**
10
10
9
8 → **Val. at 2*CS from bottom**
8
8
7
6 → **Val. at CS from bottom**
4
3
2 → **Min.**

Quantile Classes



Quantile Classes





Quantile Classes

- Advantages:
  - Each class has equal representation on the map.
  - Intuitive appeal: map readers like to be able to identify the "top 20%" or the "bottom 20%"
- Disadvantages:
  - Very irregular break points unless data have rectangular distribution.
  - Break points often seem arbitrary. Remedy this with approximate quantiles.

55

Mean & Standard Deviation

- Places break points at the ***Mean*** and at various ***Standard Deviation*** intervals above and below the mean.
- ***Mean:*** measure of central tendency.

$$Mean = (\sum X_i) / N$$

- ***Standard Deviation:*** measure of variability.

$$SD = \sqrt{\sum (X_i - Mean)^2 / N}$$

56

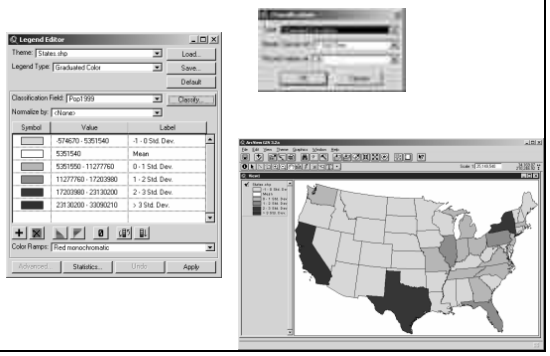
Mean & Standard Deviation

Class	Lower Limit	Upper Limit
1	Min	Mean - 2*SD
2	Mean - 2*SD	Mean - 1*SD
3	Mean - 1*SD	Mean
4	Mean	Mean + 1*SD
5	Mean + 1*SD	Mean + 2*SD
6	Mean + 2*SD	Max

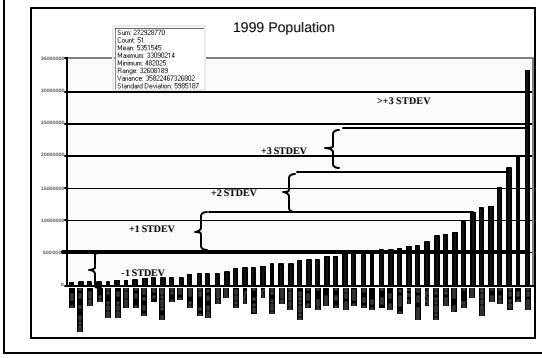
- Be careful that the calculated range doesn't lead to nonsense. For example, the Mean - 2*SD can be < Min. Clearly, the upper limit of a class may not be smaller than the lower limit. Arc View provides options additional to this one.

57

Mean & Standard Deviation



Mean & Standard Deviation



Mean & Standard Deviation

- Advantages:
  - Statistically oriented people like it.
  - Allows easier comparison of maps of variables measured in different metrics. Income and education levels.
- Disadvantages:
  - Many map readers are not familiar with the concept of the standard deviation.
  - Not good for skewed data.

60

Arithmetic Progression

- The width of each succeeding class interval is larger than the previous interval by a ***constant amount***.
- Calculating the constant amount, CW:
  - Decide on number of classes, N.
  - Calculate the range: R = Max - Min
  - Solve: R = CW + 2CW + . . . + NCW for CW

61

Arithmetic Progression

- Example CW solution:
  - Max = 160; Min = 10; R = 160 -10
  - N = 5; R = 150
  - R = CW + 2CW + 3CW + 4CW + 5CW
  - 150 = CW + 2CW + 3CW + 4CW + 5CW
  - 150 = 15CW
  - 150 / 15 = CW ; CW = 10

62

Arithmetic Progression: Class Widths

Class	Width
1	CW
2	2CW
3	3CW
4	4CW
5	5CW

63

Arithmetic Progression: Classes

Class	Lower Limit		CW		Upper Limit
1	Min	+	CW	=	Min + CW
2	Min + CW	+	2CW	=	Min + 3CW
3	Min + 3CW	+	3CW	=	Min + 6CW
4	Min + 6CW	+	4CW	=	Min + 10CW
5	Min + 10CW	+	5CW	=	Max

64

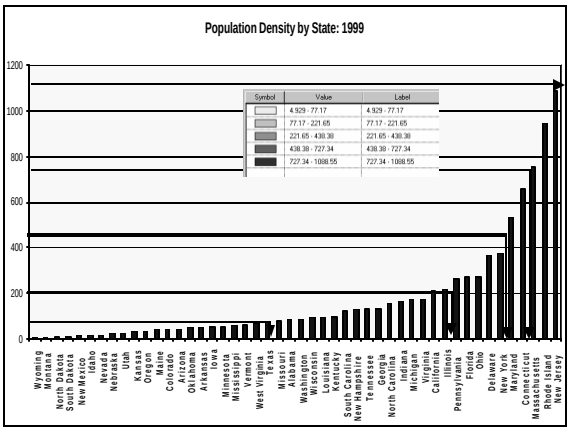
Arithmetic Progression: Class Example

- Max = 160; Min = 10; R = 150; CW = 10

Class	Lower Limit	+	CLASS* CW	=	Upper Limit
1	10	+	10	=	20
2	20	+	20	=	40
3	40	+	30	=	70
4	70	+	40	=	110
5	110	+	50	=	160

65

Arithmetic Progression



Arithmetic Progression Classes

- Advantages:
  - Uneven, but regular class breaks.
  - Tends to even out class frequencies for skewed distributions while making class widths relatively small in areas where there is high frequency.
- Disadvantages:
  - Uncommon.
  - Unequal width classes.

68

Geometric Progression Classes

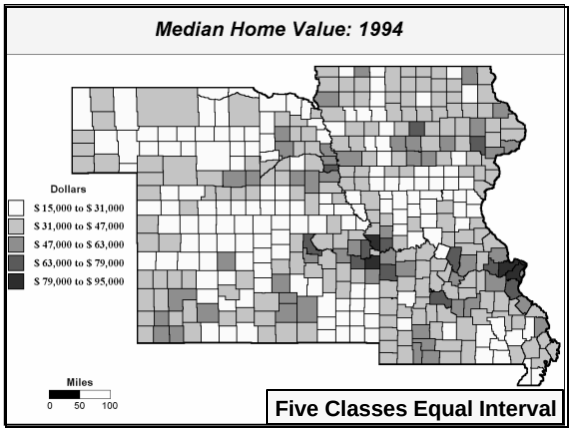
- The width of each succeeding class interval is larger than the previous interval by a exponentially varying amount.
- Calculating the BASE amount, CW:
  - Decide on number of classes, N.
  - Calculate the range: R = Max - Min
  - Solve:

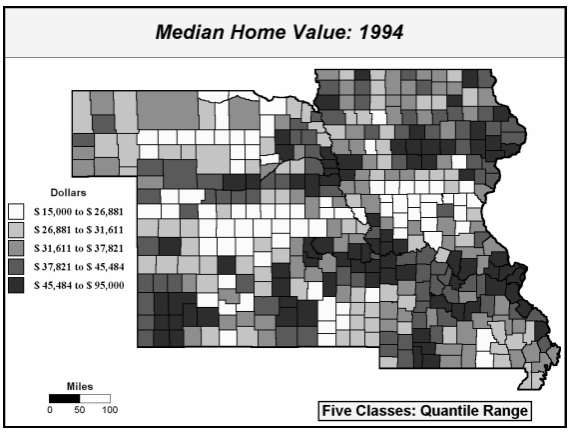
$$CW = \sqrt[N]{R}$$

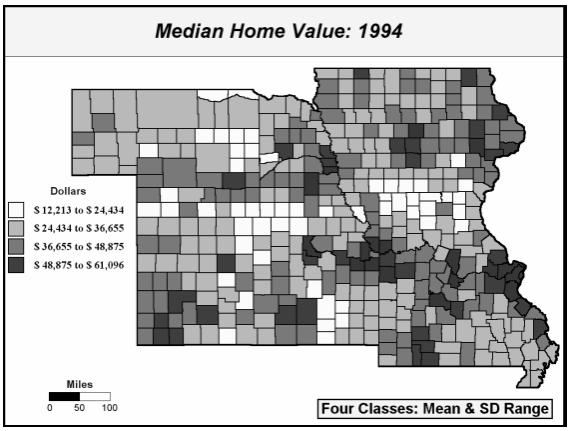
69

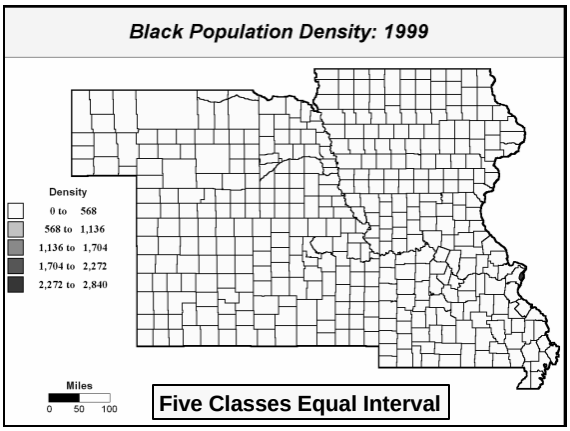
Example Maps

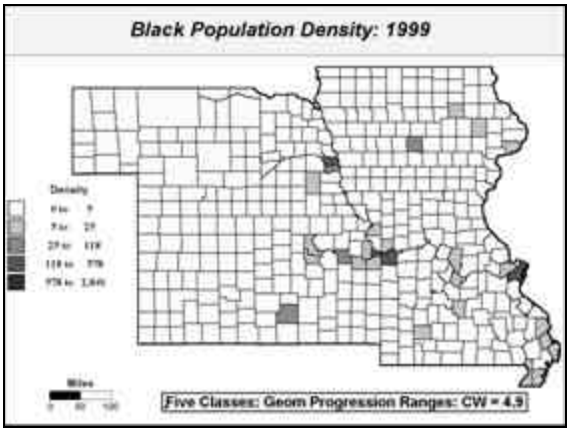
70

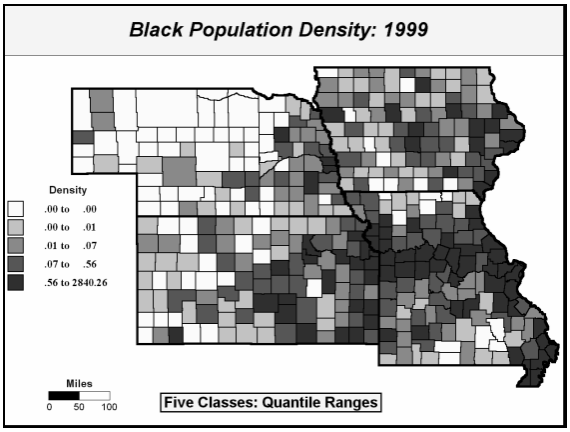





Median Home Value: 1994
Dollars
$ 12,213 to $ 24,434
$ 24,434 to $ 36,655
$ 36,655 to $ 48,875
$ 48,875 to $ 61,096
Miles
0 50 100
Four Classes: Mean & SD Range


Black Population Density: 1999
Density
0 to 568
568 to 1,136
1,136 to 1,704
1,704 to 2,272
2,272 to 2,840
Miles
0 50 100
Five Classes Equal Interval


Black Population Density: 1999
Density
Miles
Five Classes: Geom Progression Ranges: CW = 4.9


Black Population Density: 1999
Density
.00 to .00
.00 to .01
.01 to .07
.07 to .56
.56 to 2840.26
Miles
0 50 100
Five Classes: Quantile Ranges