**Topic 7: Choropleth Mapping**

Chapter 9: Chang
Chapters 4, 5, & 6: Dent

---

**Mapping Techniques Lecture Format**

- **Description**
  - Definition
  - Types/variations
  - Data characteristics
    - Type of data
      - Raw or Derived
    - Spatial characteristics
      - Discrete or Continuous
- **Design Considerations**
  - Projection
  - Legend
  - Symbology
    - Classification
    - Colour scheme
  - Scale
  - Other

---

**What is Choropleth Mapping?**

- Uses distinct
  1. Administrative or natural
  2. Enumeration area mapping
  3. ... to show changes in value, by areal unit

---
Two Types of Choropleth Maps

1. Conventional or Simple

2. Classless, Unclassed, or Tonal

Types of Spatial Data

Choice of mapping technique often determined by type of data

Both spatial properties
   • Discrete vs. Continuous

And aspatial properties
   • Raw vs. Derived

Spatial Data Characteristics

Discrete Data (review)
   • Values that occur at:
     • a point
     • a line
     • or polygon

   • No data between features
   • Anthropogenic phenomena
Spatial Data Characteristics

- Discrete areal units, very common
  - Typically statistical or administrative areas
  - Data are aggregated for entire areal unit
  - May be:
    - Totals
    - Averages
    - Rates
    - Proportions
  - In reality may not be uniform or actually occur at any point

Spatial Data Characteristics

- Continuous (review)
  - Occur across entire area
  - Sampled at specific locations (points)
  - Vary continuously
  - Naturally occurring phenomena

Spatial Data Requirements

- Choropleth technique requires data aggregated by discrete areal unit
- What about other types of spatial data?
Two Aspatial Data Types

- Specific techniques are used
  - Two types
    1. Raw Values
    2. Derived Data
      - ratios, proportions, percentages, rates
      - frequently normalized by area or population
  - Choropleth techniques uses derived data

Types of Discrete Areal Units

- Natural Areal Units
  - Boundaries of natural discrete phenomena
- Artificial Areal Units
  - Arbitrary imposed boundaries

Modifiable Areal Unit Problem

- The regionalization space
- Data collected using arbitrary boundaries
- What would happen if we imposed a different set of boundaries?
Modifiable Areal Unit Problem

Considerations: Number & Size of Areal Units
- Areal units often hierarchically nested
- Data available at multiple levels
- So we have a choice

Considerations: Map Scale
- Smallest areal unit must be visible
- What determines scale?
  - 
  - 
  -
Considerations: Classification Technique

- **Significantly** impacts message
- More than one version can be presented; not common
- Should use **most appropriate method**; not one that produces desired effect
- Statement indicating technique should be included

Data Classification

- Purpose is to:
  1. 
  2. 
  3. 

Classification Schemes

- Two classifications of classification schemes
  1. Exogenous - derived externally
  2. Arbitrary
  3. Idiographic - unique characteristics
  4. Serial - a systematic progression
Classification Schemes

- Second classification based on intervals

1. Constant Intervals

2. Variable Intervals

Equal Steps

- Each class represents equal proportion of the range of data values
- Constant interval scheme
- Procedure:

  Data range \( R = H - L \)

  Interval \( I = R / n \) (# classes)

  Class boundaries are then determined by:

  \[ L, L + (1 \times I), L + (2 \times I), \ldots, L + (n \times I) \]

  Lower Boundary  Upper Boundary

Equal Steps

<table>
<thead>
<tr>
<th>Province</th>
<th>Population 1991</th>
<th>Area (km²)</th>
<th>Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.W.T.</td>
<td>64,462.00</td>
<td>3,240,390.49</td>
<td>0.02</td>
</tr>
<tr>
<td>Yukon Territory</td>
<td>30,766.00</td>
<td>579,670.72</td>
<td>0.00</td>
</tr>
<tr>
<td>New Brunswick</td>
<td>551,762.00</td>
<td>711,646.56</td>
<td>0.78</td>
</tr>
<tr>
<td>Saskatchewen</td>
<td>790,982.00</td>
<td>651,110.47</td>
<td>1.24</td>
</tr>
<tr>
<td>Manitoba</td>
<td>1,133,860.00</td>
<td>647,740.95</td>
<td>2.13</td>
</tr>
<tr>
<td>B.C./Saskatchewan</td>
<td>3,726,280.00</td>
<td>582,778.30</td>
<td>6.37</td>
</tr>
<tr>
<td>Alberta</td>
<td>2,050,630.00</td>
<td>660,282.30</td>
<td>3.13</td>
</tr>
<tr>
<td>Quebec</td>
<td>7,319,872.00</td>
<td>1,357,817.73</td>
<td>5.38</td>
</tr>
<tr>
<td>New Brunswick</td>
<td>738,113.00</td>
<td>711,646.23</td>
<td>10.31</td>
</tr>
<tr>
<td>Ontario</td>
<td>12,703,173.00</td>
<td>992,713.78</td>
<td>13.83</td>
</tr>
<tr>
<td>Nova Scotia</td>
<td>509,292.00</td>
<td>52,840.83</td>
<td>9.72</td>
</tr>
<tr>
<td>Prince Edward Island</td>
<td>154,550.00</td>
<td>5,449.16</td>
<td>28.56</td>
</tr>
</tbody>
</table>

R = 23.77 - 0.02 = 23.75
L = 23.75 / 5 = 4.75
Class Boundaries are:
0.02, 4.77, 9.52, 14.27, 19.02, 23.77
Note: No overlapping classes
Equal Steps

- Most appropriate when?
  - 
  -
- Neither is a common occurrence
- Accentuates outliers

![Histogram](image1)

Equal Steps

![Histogram](image2)

![Map](image3)
Standard Deviations

- Boundaries based on mean and SD
- Each class = equal proportion of total deviation
  - constant interval scheme
- Most appropriate when?
Standard Deviations

Procedure:
- mean value \( m = \frac{\sum x}{n} \)
- \( SD = \left[ \frac{\sum (x - m)^2}{n-1} \right]^{1/2} \)

- Class boundaries are then determined by:
  \( m + (1 \times SD) \) and \( m - (1 \times SD) \)

- Usually up to 6 classes
  - 3 above mean
  - and 3 below mean

---

### Standard Deviations

<table>
<thead>
<tr>
<th>Province</th>
<th>Population 1996</th>
<th>Acres/km²</th>
<th>Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alberta</td>
<td>3,720,000</td>
<td>316,000</td>
<td>11.72</td>
</tr>
<tr>
<td>British Columbia</td>
<td>4,170,000</td>
<td>582,000</td>
<td>7.17</td>
</tr>
<tr>
<td>Manitoba</td>
<td>996,000</td>
<td>674,000</td>
<td>1,473</td>
</tr>
<tr>
<td>New Brunswick</td>
<td>728,000</td>
<td>77,500</td>
<td>9.42</td>
</tr>
<tr>
<td>Newfoundland and Labrador</td>
<td>509,200</td>
<td>409,000</td>
<td>1.25</td>
</tr>
<tr>
<td>Nova Scotia</td>
<td>1,344,000</td>
<td>25,000</td>
<td>53.77</td>
</tr>
<tr>
<td>Ontario</td>
<td>13,710,000</td>
<td>1,757,000</td>
<td>7.77</td>
</tr>
<tr>
<td>Quebec</td>
<td>7,138,000</td>
<td>1,179,000</td>
<td>6.02</td>
</tr>
<tr>
<td>Yukon</td>
<td>30,760</td>
<td>531,000</td>
<td>0.06</td>
</tr>
<tr>
<td>N.W.T.</td>
<td>39,170</td>
<td>271,000</td>
<td>0.14</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>1,020,000</td>
<td>275,000</td>
<td>3.74</td>
</tr>
</tbody>
</table>

Mean = 82.01/12 = 6.83

SD = 7.5

Note: No overlapping classes

Six Classes

\(<1 \ SD \ 0 - 6.82\>

\(>1 \ SD \ 6.83 - 14.32\>

\(>2 \ SD \ 14.33 - 21.15\>

\(>3 \ SD \ 21.16 - 27.99\>

---

### Standard Deviations

![Histogram and Frequency Chart]

Population Density

Frequency

Cumulative %
Standard Deviations

- Creates a more “even” looking distribution
- May present interpretation issues
Geometric Intervals

- Class limits based on arithmetic or geometric properties of data
- Most appropriate when?
- A variable interval technique
- Less common
Quantiles

- Boundaries selected so that **same number** of areal units in each class
- Intervals not constant

**Procedure:**
1. arrange all values in ascending order
2. determine number of obs in each class (K) by:
   \[ K = \frac{\text{# obs}}{\text{# classes}} \]
3. Starting at the lowest value, place an equal number of observations in each class
4. Class limit is mean value between adjacent observations in different classes

**Example:**
- K = 12 obs / 4 classes = 3 obs/class
- Four Classes:
  - 0.02 – 1.60
  - 1.61 – 4.19
  - 4.20 – 11.01
  - 11.02 – 23.77

Still, no overlapping classes
Quantiles

- Produces an “even” looking map
  - A sense of diversity when there is little
  - Effectively masks outliers

Natural Breaks (Manual)

- Based on visual inspection of data using
- Boundaries where natural groupings occur
- Number of classes determined by breaks
- Subjective technique, but can be effective
Jenk’s Optimization Method

- Based on goodness of variance fit (GVF)
- Maximizes:
  - between class heterogeneity
  - within class homogeneity
- You pick number of classes
### Jenk’s Optimization Method

#### Natural Breaks – 5 classes

<table>
<thead>
<tr>
<th>Province</th>
<th>Population</th>
<th>Area (km²)</th>
<th>Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northwest Territories</td>
<td>64,442.99</td>
<td>2,240,285.45</td>
<td>0.03</td>
</tr>
<tr>
<td>Yukon Territory</td>
<td>36,769.00</td>
<td>311,645.52</td>
<td>0.11</td>
</tr>
<tr>
<td>Nunavut</td>
<td>161,740.93</td>
<td>271,636.63</td>
<td>0.60</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>360,257.96</td>
<td>575,115.47</td>
<td>0.62</td>
</tr>
<tr>
<td>Manitoba</td>
<td>1,153,640.90</td>
<td>547,703.95</td>
<td>2.12</td>
</tr>
<tr>
<td>Alberta</td>
<td>2,724,690.90</td>
<td>863,677.99</td>
<td>3.17</td>
</tr>
<tr>
<td>British Columbia</td>
<td>2,060,630.90</td>
<td>630,236.93</td>
<td>3.26</td>
</tr>
<tr>
<td>Quebec</td>
<td>7,136,797.99</td>
<td>1,372,541.73</td>
<td>5.20</td>
</tr>
<tr>
<td>New Brunswick</td>
<td>738,713.93</td>
<td>75,589.23</td>
<td>9.76</td>
</tr>
<tr>
<td>Ontario</td>
<td>16,750,573.90</td>
<td>190,730.79</td>
<td>11.93</td>
</tr>
<tr>
<td>Nova Scotia</td>
<td>499,230.93</td>
<td>52,440.63</td>
<td>9.72</td>
</tr>
<tr>
<td>Prince Edward Island</td>
<td>434,617.93</td>
<td>5,660.38</td>
<td>7.67</td>
</tr>
</tbody>
</table>

GVF = (SDAM - SDCM) / SDAM

Where:
- SDAM = sum of squared deviations from array mean
- SDCM = sum of squared deviations from class means

- When SDCM is lowest, then GVF will be closest to 1
- This is the best set of five classes

### Jenk’s Optimization Method

\[ \text{SDAM} = \sum (x - \bar{x})^2 \]

Where:
- \( x \) = the array mean = 6.83
- \( \bar{x} \) = each data value

\[ \text{SDCM} = \sum (x - \bar{X})^2 \]

Where:
- \( X \) = each data value
Jenk's Optimization Method

\[ \text{SDCM} = \sum (z_c - x_i)^2 \]

Where:
\[ z_c = \text{the class mean} \]
\[ x_i = \text{each data value} \]

\[ \sum (z_c - x_i)^2 \]

\[ \sum (z_c - x_i)^2 \]

\[ GVF = 616.95 - 5.42 / 616.95 = 0.99 \]

Jenk's Optimization Method

<table>
<thead>
<tr>
<th>Province</th>
<th>Population 1996</th>
<th>Area (km²)</th>
<th>Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northwest Territories</td>
<td>64,422</td>
<td>3,286,896</td>
<td>0.02</td>
</tr>
<tr>
<td>Yukon</td>
<td>30,760</td>
<td>3,286,896</td>
<td>0.09</td>
</tr>
<tr>
<td>Newfoundland</td>
<td>551,762</td>
<td>97,608,517</td>
<td>5.60</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>999,237</td>
<td>97,608,517</td>
<td>10.20</td>
</tr>
<tr>
<td>Manitoba</td>
<td>1,113,698</td>
<td>647,759</td>
<td>17.20</td>
</tr>
<tr>
<td>British Columbia</td>
<td>3,724,820</td>
<td>1,257,941</td>
<td>29.47</td>
</tr>
<tr>
<td>Alberta</td>
<td>2,998,709</td>
<td>0.667,941</td>
<td>43.30</td>
</tr>
<tr>
<td>Quebec</td>
<td>7,148,166</td>
<td>7,196,611</td>
<td>10.01</td>
</tr>
<tr>
<td>New Brunswick</td>
<td>1,733,273</td>
<td>7,196,611</td>
<td>24.02</td>
</tr>
<tr>
<td>Ontario</td>
<td>10,753,573</td>
<td>1,751,720</td>
<td>1,117.73</td>
</tr>
<tr>
<td>Nova Scotia</td>
<td>269,260</td>
<td>52,661,831</td>
<td>17.20</td>
</tr>
<tr>
<td>Prince Edward Island</td>
<td>134,597</td>
<td>5,688,380</td>
<td>23.77</td>
</tr>
</tbody>
</table>

\[ z_c = 1.06 \]
\[ z_c = 4.55 \]
\[ z_c = 11.02 \]
\[ z_c = 17.20 \]
\[ z_c = 23.77 \]
Jenk's Optimization Method

<table>
<thead>
<tr>
<th>Province</th>
<th>Population (16+)</th>
<th>Area (sq mi)</th>
<th>Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northwest Territories</td>
<td>64,432,061</td>
<td>2,289,388.61</td>
<td>0.07</td>
</tr>
<tr>
<td>Yukon Territory</td>
<td>30,785,012</td>
<td>70,951.72</td>
<td>0.43</td>
</tr>
<tr>
<td>Nunavut</td>
<td>47,785,012</td>
<td>1,067,645.21</td>
<td>0.01</td>
</tr>
<tr>
<td>Yukon Territory</td>
<td>4,721,012</td>
<td>1,113,600.21</td>
<td>1.00</td>
</tr>
<tr>
<td>British Columbia</td>
<td>2,106,012</td>
<td>882,577.59</td>
<td>2.45</td>
</tr>
<tr>
<td>Alberta</td>
<td>2,096,012</td>
<td>905,212.64</td>
<td>2.35</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>4,498,012</td>
<td>2,925,941.73</td>
<td>0.02</td>
</tr>
<tr>
<td>New Brunswick</td>
<td>1,238,012</td>
<td>71,599.23</td>
<td>10.34</td>
</tr>
<tr>
<td>Ontario</td>
<td>11,703,012</td>
<td>1,913,725.75</td>
<td>6.05</td>
</tr>
<tr>
<td>Nunavut</td>
<td>1,050,012</td>
<td>564,681.83</td>
<td>1.87</td>
</tr>
<tr>
<td>Porth Gwin and Island</td>
<td>134,587.01</td>
<td>5,609.38</td>
<td>23.77</td>
</tr>
</tbody>
</table>

Quantiles with 4 classes
GVF = 616.95 - 99.87 / 616.95
= 0.84

Considerations: Legend Design

- Classes are typically range graded
- Refers to use of classes with continuous intervals
  - E.g. 1-10, 11-20, 21-30, 31-40,
- But data may be non-continuous
- Interpretation errors
Legend Design

Effect of range grading
Non-continuous data

Legend Design/Ancillary Data

- Guidelines for legend boxes
  - 2/3 proportion
  - or irregular shapes

- Appropriate ancillary data include:
  - histogram w/ cumulative % curve
  - indication of classification technique

Considerations: Symbol Selection

- Symbols indicate relative change in value
- Achieved by varying symbol:
  - Arrangement
  - Texture
  - Orientation
  - Colour saturation/chroma
  - Colour value/intensity
  - Colour hue
Considerations: Map Projection

- Proportion of mapped area represented by different symbols affects interpretation
- Should use an equivalent projection