PDHonline Course L155G (5 PDH)

Data Models and Data processing in GIS

Instructor: Steve Ramroop, Ph.D.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

An Approved Continuing Education Provider

Slide 1

Lecture 4 Content

❑ **Geographic Information Systems (GIS)**

Data Models, Data Structure and Data Management

(Continued)

This lecture is the continuation of the GIS topic identified in the course description which is Data Models, Data Structure and Data Management.

Slide 2

Lecture Content

- **Spatial Data Models**
  - **How to create raster data?**
  - **How to store raster data?**
    - ➢ **full run encoding**
    - ➢ **run length encoding**
    - ➢ **quadtree representation**
  - **comparison of raster and vector data**

This slide show the content covered in this lecture. The main detail of this lecture is the characteristics of the raster data structure; its creation; and its types of data storage methods. The conclusion of the lecture gives a comparison of the vector and raster data structure.

Slide 3

b)<u>Raster Model Representation</u>

– The raster model represents data as a regular grid of squares cells
– the location of each cell or “pixel” (picture element) is defined by its row and column numbers
– value assigned to the pixel is the value of the attribute it represents

This is the second data structure used in GIS which is called the raster data structure. The graphic primitive of this structure is a pixel. The pixel is repeated and used to represent lines and polygons. A single pixel will be used as a point. The size of the pixel indicates the quality of the raster structure data set. Small sizes are a higher quality than larger sized pixels. For example: 1cm pixels are at a higher quality than 20m pixels.

Slide 4

- the raster data structure covers the entire area into a regular grid of cells in a specific sequence:
  - row by row
  - each cell contain single value
  - space filling
- TELLS WHAT OCCURS EVERYWHERE
- raster database can consist of many layers (e.g. soil type, elevation, land use, land cover, etc.)
- The diagram on the next slide gives an idea of the representation of the raster data structure

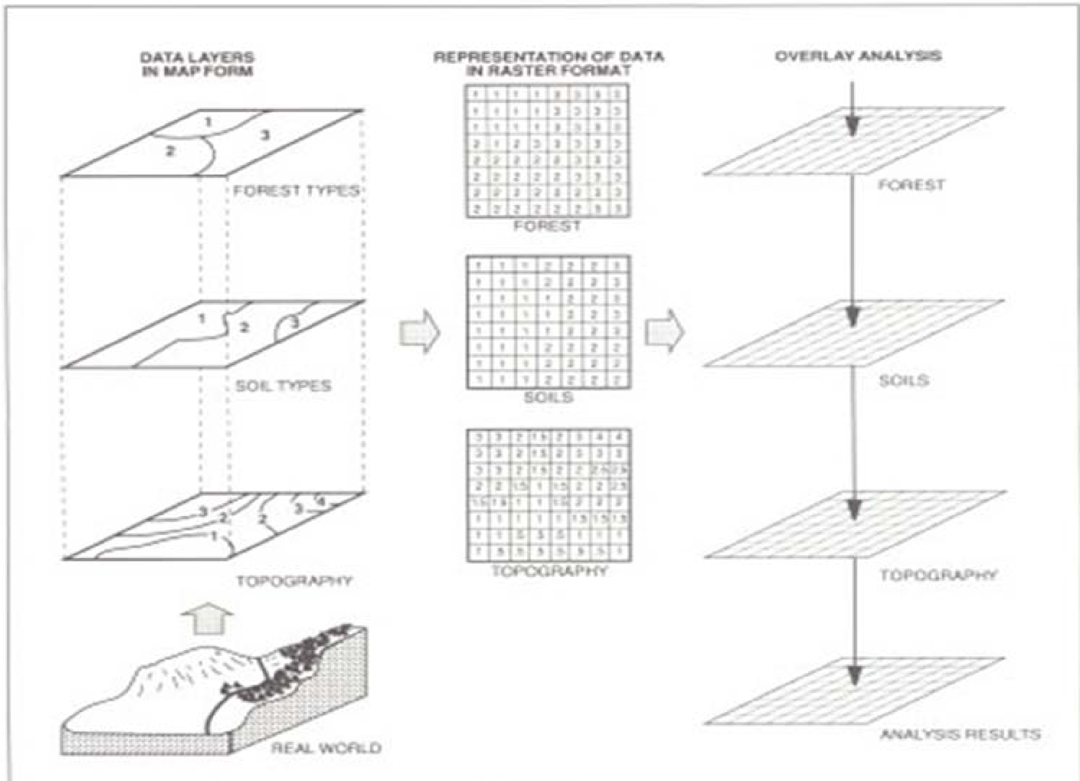
L155 - GIS Data Models and Data Processing
Lecture 4
Dr. Steve Ramroop

4

This slide shows some of the common characteristics of the raster data structure. In this structure there is a value for every pixel….even if the pixel has a value of 0. When compared to the vector data structure is different. For the vector data structure, areas of no value do not have a specific value. The raster data structure is space filling and the vector data structure is not. The phrase “Tells what occurs everywhere” refers that if the phenomena for a given data set does not occur at that location, a null value is stored. Therefore, for a given area the raster data structure will values for each pixel.

Slide 5

Overlay Analysis using Raster Data files



L155 - GIS Data Models and Data Processing
Lecture 4
Dr. Steve Ramroop

This is an example showing the comparison of the vector data structure with that of the raster data structure.

Slide 6

How to create a Raster Data Structure?

1) Entering cell values using a word processor, database or spreadsheet program for each layer cell by cell
   – tedious and time consuming

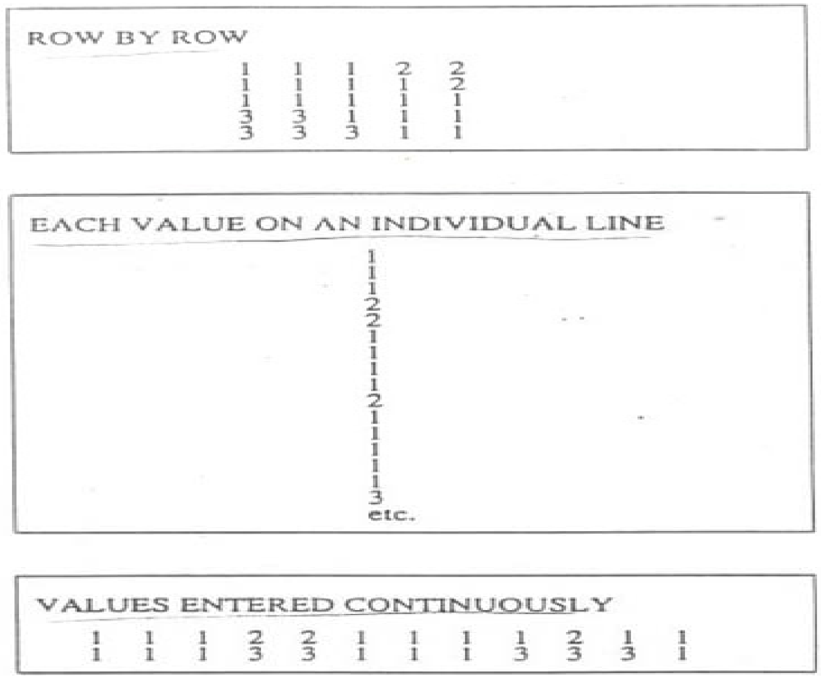
2) Digitize existing maps using a digitizing table in vector formats and then convert the softcopy into raster formats

3) Scanning the hardcopy map
   – Using a scanner, hardcopy maps are converted into softcopy maps

4) Some data (e.g. Remote Sensing data) are directly captured and stored in raster formats
   – requires additional image pre-processing before usage

This slide shows four methods which are self-explanatory regarding the creation of raster data structure. Note that the conversion from vector to raster is called rasterization while the reverse from raster to vector is called vectorization.

Slide 7



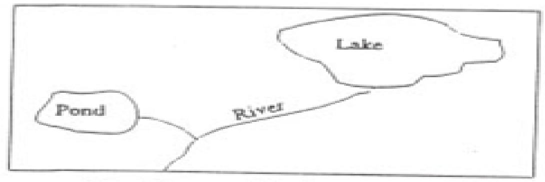
Typical ASCII file formats used for importing raster data

L155 - GIS Data Models and Data Processing
Lecture 4
Dr. Steve Ramroop

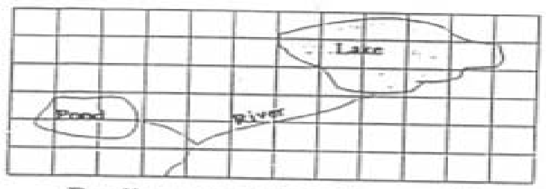
This slide show the typical structure used to create raster data using ASCII file formats by typing the pixel values using a word processor or text editor.

Slide 8

Creating raster data



Reality - Hydrography



Reality overlaid with a grid

0	0	0	0	0	0	0	1	1	1	0	0
0	0	0	0	0	0	1	1	1	1	1	0
0	0	0	0	0	0	0	1	1	1	0	0
0	1	1	0	0	0	2	2	0	0	0	0
0	1	1	2	2	2	0	0	0	0	0	0
0	0	0	2	0	0	0	0	0	0	0	0

0 = No Water Feature
1 = Water Body
2 = River

Resulting raster

This slide shows another example of converting from the vector data structure into the raster data structure. Note that if the area covered for a given cell is more than half then the entire cell is considered in the final raster data structure.

Slide 9

Cell values in a Raster data layer

- The cell values depend upon both:
  - the reality being coded (eg. Trees maybe coded as numerical vales or alphabetic values); and
  - the GIS software being used which may have restrictions on the type of cell values allowed
- Different systems allow different types of values. For example values can be:
  - Whole number (integer)
  - Real (decimal) values
  - Alphabetic (character) values
- Integer values often act as code numbers which point to names in an associated table or legend (eg. 1 can represent roads; 2 can represent buildings; and such like)

This slide presents a description of the values of pixels which are used to describe the attributes of the raster data structure. Always try to use values which can be directly understood by the GIS user and is indicative of the phenomena being mapped.

Slide 10

Raster data values

0	0	1	1	1	2	3
0	1	1	1	2	2	3
0	0	1	1	1	2	2
0	0	0	1	1	1	2
0	0	0	0	0	0	1

INTEGER VALUES

e.g. soil class

2.1	2.3	2.4	2.7	2.9	3.2	3.4
1.9	2.1	2.4	2.6	3.0	3.3	3.4
1.8	2.0	2.3	2.5	2.8	3.1	3.2
1.8	1.9	1.9	2.3	2.5	2.8	3.0
1.8	1.8	1.8	1.9	2.1	2.3	2.7

REAL (DECIMAL) VALUES

e.g elevation

a	a	b	b	c	c	a
a	b	b	b	c	c	a
a	b	b	b	b	c	c
a	a	b	b	b	c	c
a	a	a	b	c	c	c

NON-NUMERIC VALUES

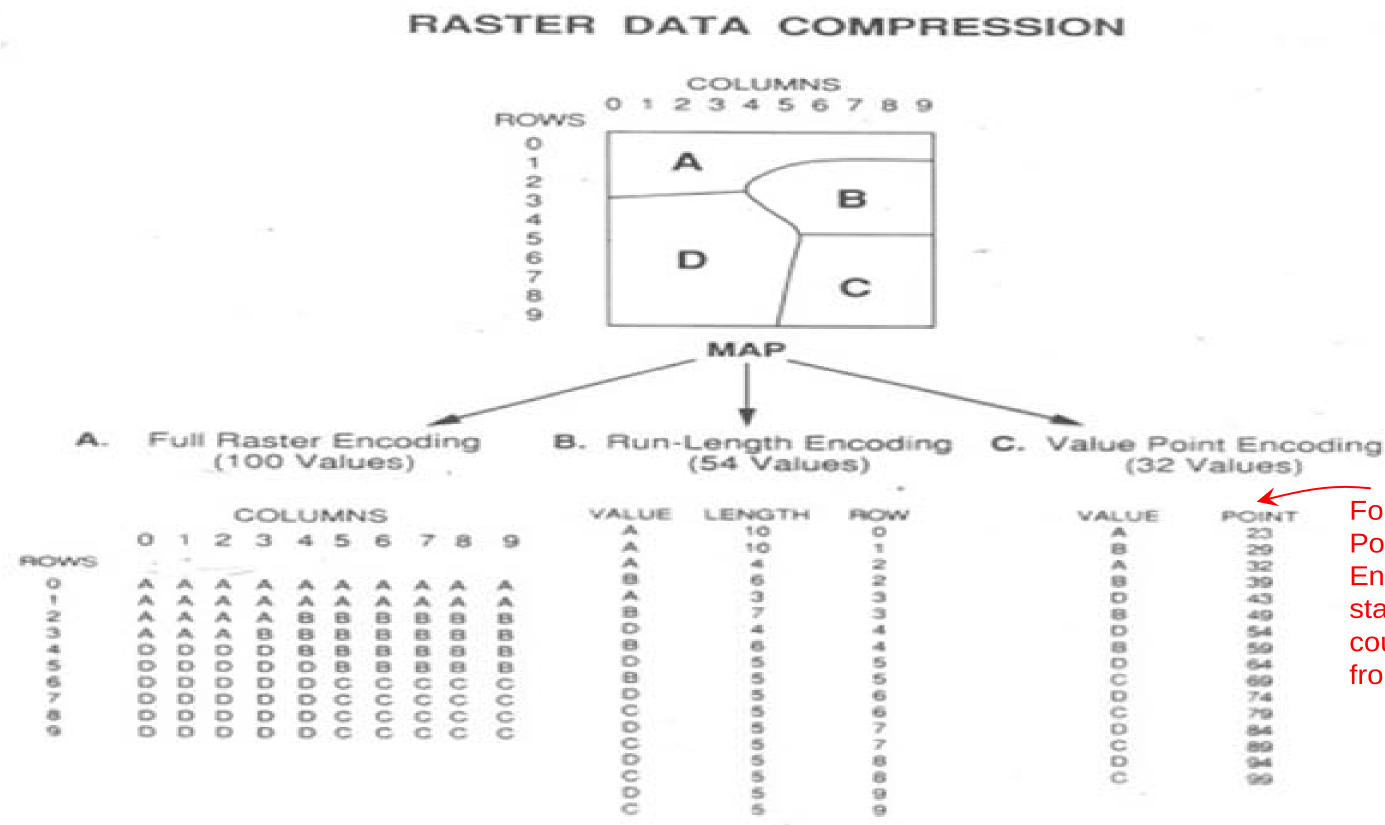
e.g. vegetation class

a – [illegible]
b – [illegible]
c – [illegible]

This slide shows three types of coded raster data representing three different phenomena from reality.

Slide 11

Standard Run length raster encoding



For Value Point Encoding start counting from '0'

This slide summarizes three methods used to store raster data structures. Some general comments of slide 11 are as follows:

- The first method of storing Raster data structure is called **“Full Raster Encoding”**.
- The second method of storing raster data structure is called **“Run Length Encoding”**. There are two categories of “Run Length Encoding” which are:
  - **“Standard Run Encoding”** and
  - **“Value Point Encoding”**

From this diagram you must note how the data is stored. The “Full Raster Encoding”, method stores 100 values; the “Standard Run Encoding” method stores 54 values; while the “Value Point Encoding” method stores 32 values. The raster data is being compressed from its original “Full Raster Encoding”.

Note that the “Value Point Encoding” method starts counting from the number ‘0’.

Details into each method are discussed in the slides that follow.

Slide 12

❑ How to store a Raster data?

There are three methods used to store raster data:

1) **Full Raster encoding**
2) **Run - Length Encoding**
3) **Quadtree Representation**

1) Full Raster Encoding

- **each cell having its particular code is stored individually**
- **problem of storage space**
- **processing speed is reduced**
- **redundancy in database**

This slide identifies three methods of raster data structure storage. Characteristics of the Full Raster Encoding method are listed.

Slide 13

2) Run - Length Encoding

- **adjacent cells along a row that have the same value are treated as a group termed a *“run”***
- **the pixel value is stored once, together with information about the size and location of the run**

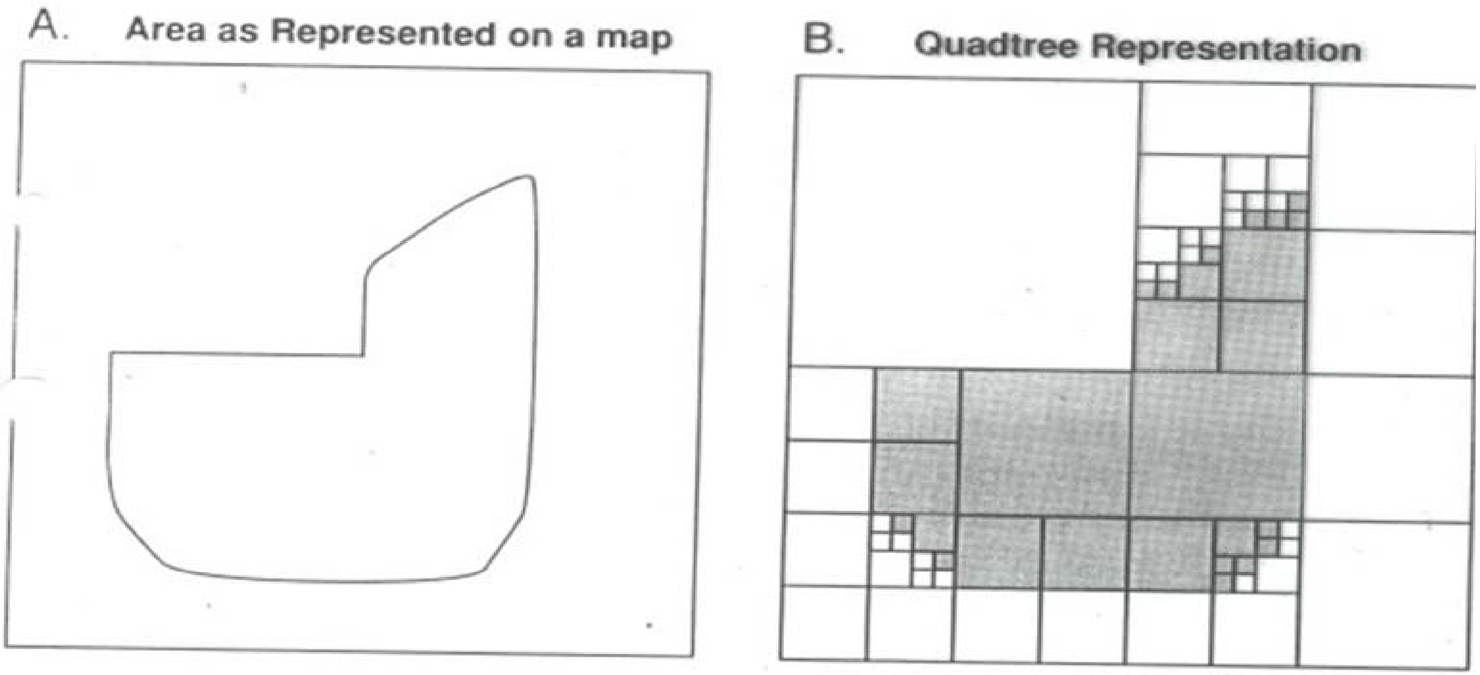
This slide presents some characteristics of the “Run Length Encoding” method of storing raster data structures. The two categories of the “Run Length Encoding” method are presented in the next slide.

Slide 14

- There are two categories of run - length encoding:
  - a) Standard Run - Length Encoding
    - the value of the attribute, the number of the cells in the run and the row number are recorded in a file
  - b) Value Point Encoding
    - cells are assigned position numbers starting in the upper left corner of the image, proceeding from left to right and from top to bottom
    - position number at the end of each run is stored in the “POINT” column while the value for each cell is stored in the “VALUE” column in a file
    - Start counting from the first entry with ‘0’

These are the two methods discussed earlier (method B and C on slide 11). The understanding for how the data is stored is described on this slide. Value Point Encoding method makes use of least storage space.

3) Quadtree Representation



The third method of storing raster data structure is presented here. It makes use of quarter section sub-divisions. The left figure shows the study area of interest while the right figure shows the Quadtree Representation of the study area of interest.

The Quadtree Representation has multiple sized grid cells. The more detail and irregular is the dataset, the more quarter section sub-divisions there will be.

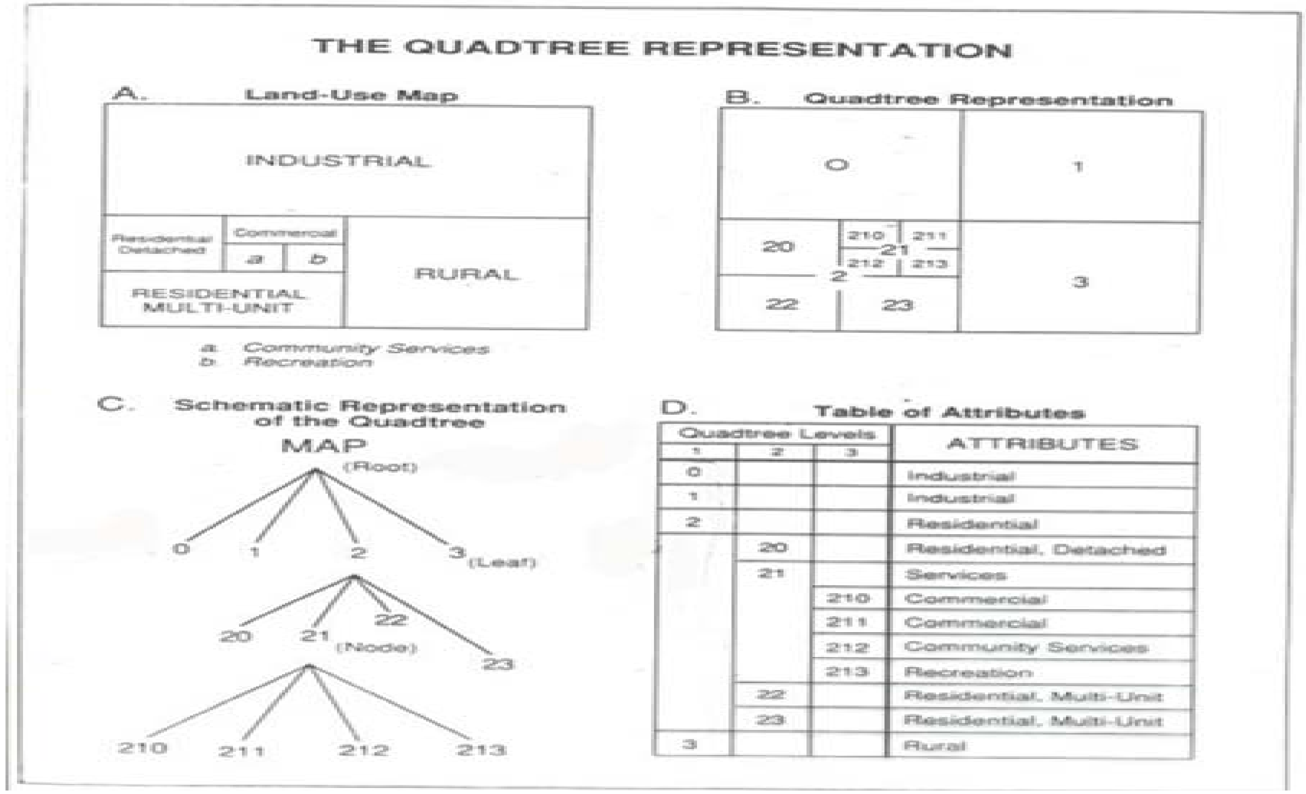
Slide 16

- **provides a more compact raster representation by using a *<u>variable sized grid cell</u>***
- **finer detailed areas will require the grid cell to be further sub-divided**
- **finer resolution (small cells) is used for areas of high spatial variability**
- **The following slide shows an example of a quadtree representation**

This slide presents some characteristics of the Quadtree Representation.

Slide 17

Example of a raster Quadtree Representation



Quadtree Levels 1	2	3	ATTRIBUTES
0			Industrial
1			Industrial
2			Residential
	20		Residential, Detached
	21		Services
		210	Commercial
		211	Commercial
		212	Community Services
		213	Recreation
	22		Residential, Multi-Unit
	23		Residential, Multi-Unit
3			Rural

In this example the figure shows four figures labeled A, B, C, and D. Figure A shows the Land-Use Map to me mapped using the raster Quadtree Representation. The Land-Use Map has a classification of Land-Use for example: Industrial, Rural, Residential, and such like. The classes are regular quadrilateral shapes.

Figure B divides the entire Land-Use Map into quarter section sub-divisions and at the same time numbering the sections. The first Quadtree level 1 divides the study area into four quarters that are numbered 0, 1, 2, and 3 as shown in Figure B. The quarters 0, 1 and 3 requires no further section sub-divisions however quarter 2 requires further section sub-divisions into the second Quadtree level 2. The numbering of the second Quadtree level 2 is 20, 21, 22 and 23. The section sub-divisions 20, 22 and 23 do not require further section sub-divisions however quarter 21 requires further section sub-divisions into 210, 211, 212 and 213.

The Schematics Presentation of the Quadtree is shown in Figure C.

Figure D shows the Table of attributes that stores the Land-Use classification into a table and its associated Quadtree Levels.

- major disadvantage is the time it takes to create and modify the quadtree
- requires more processing time to generate the quadtree with its indexes and tables
- if the data is fairly homogeneous then quadtrees provide efficient storage
- fewer the classes larger the clumps greater the degree of compression and more efficient is the quadtree structure
- best utilized when the need for updating is not frequent

This slide presents some pros and cons of the quadtree representation. This representation is best used for data sets that are regular quadrilateral shapes.

Slide 19

Table 6.1 Comparison of Raster and Vector Data Models.

RASTER MODEL	VECTOR MODEL
Advantages:	**Advantages:**
1. It is a simple data structure.	1. It provides a more compact data structure than the raster model.
2. Overlay operations are easily and efficiently implemented.	2. It provides efficient encoding of topology, and, as a result, more efficient implementation of operations that require topological information, such as network analysis.
3. High spatial variability is efficiently represented in a raster format.	3. The vector model is better suited to supporting graphics that closely approximate hand-drawn maps.
4. The raster format is more or less required for efficient manipulation and enhancement of digital images.	
Disadvantages:	**Disadvantages:**
1. The raster data structure is less compact. Data compression techniques can often overcome this problem.	1. It is a more complex data structure than a simple raster.
2. Topological relationships are more difficult to represent.	2. Overlay operations are more difficult to implement.
3. The output of graphics is less aesthetically pleasing because boundaries tend to have a blocky appearance rather than the smooth lines of hand-drawn maps. This can be overcome by using a very large number of cells, but may result in unacceptably large files.	3. The representation of high spatial variability is inefficient.
	4. Manipulation and enhancement of digital images cannot be effectively done in the vector domain.

This slide identifies some of the pros and cons about raster and vector data structures. They are all self explanatory.

Slide 20

… The End …