Week 2: Data Models

GEOG 011

Dr. Evan Lue

Pasadena City College

Spring 2015

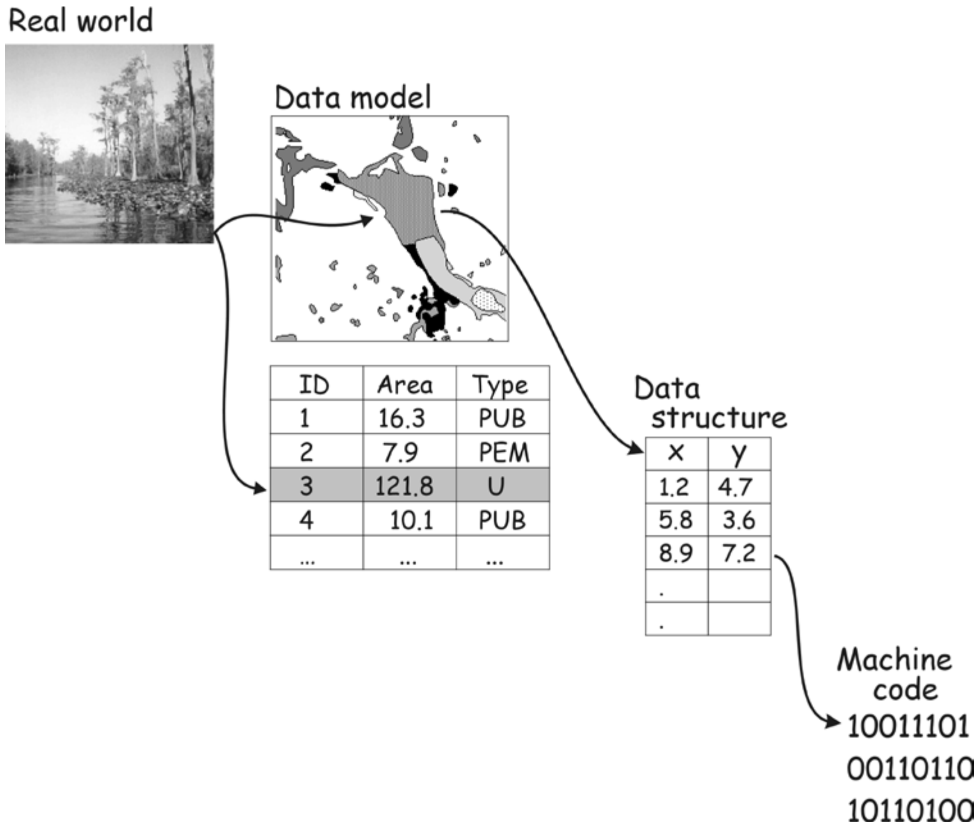
Introduction

Entities

Spatial features
(sometimes called cartographic objects)

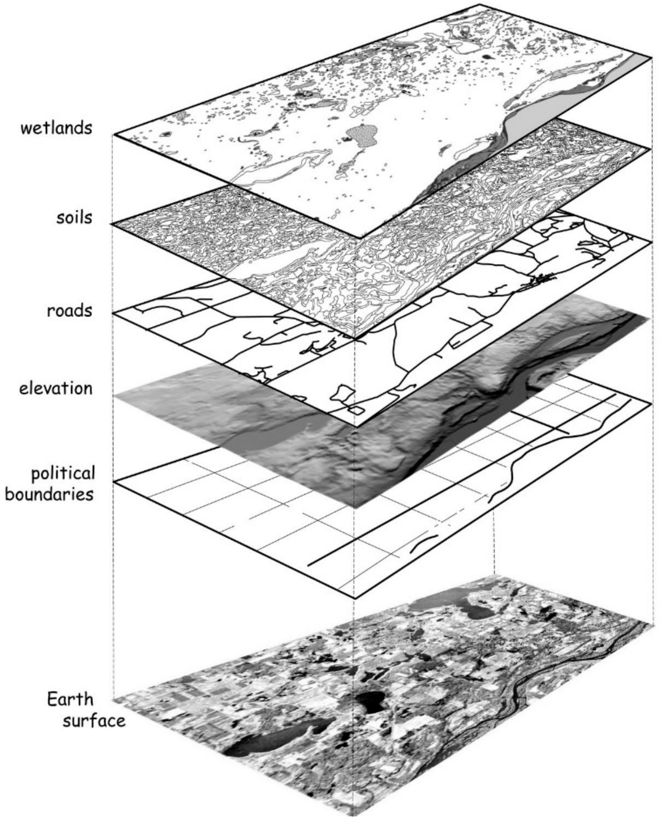
Bolstad 2012, Fig 2-1

Spatial Data Models



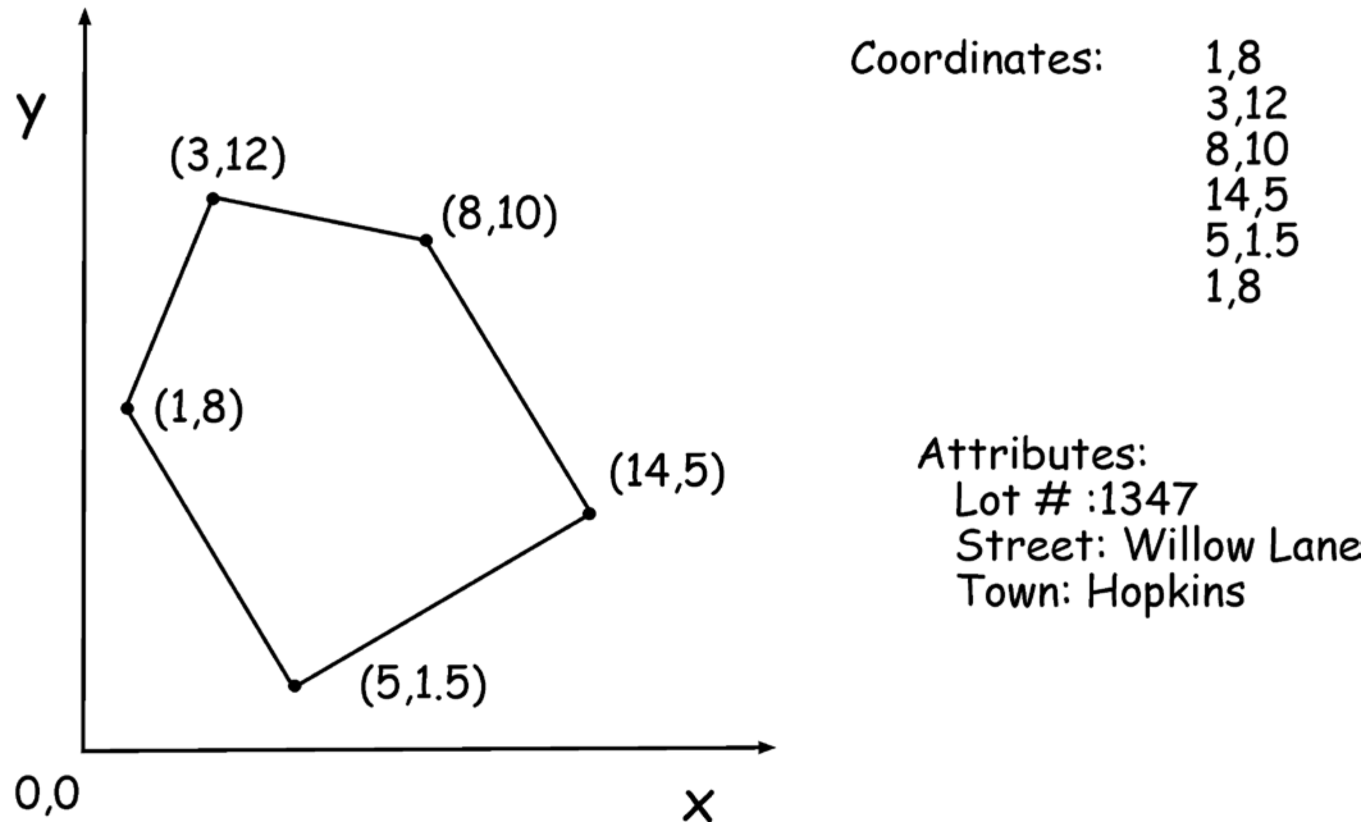
Bolstad 2012, Fig 2-2

Thematic Layers



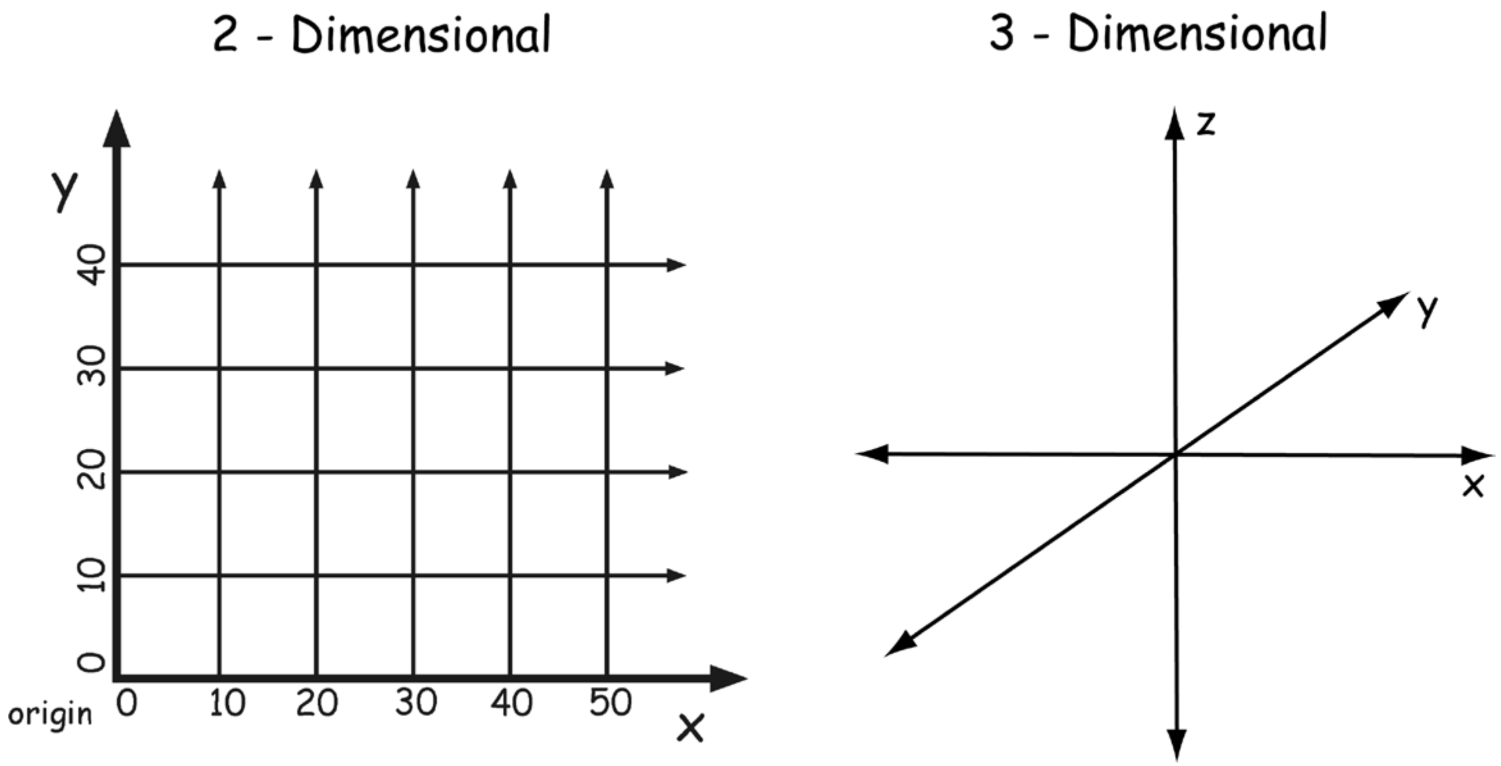
Bolstad 2012, Fig 2-4

Coordinates



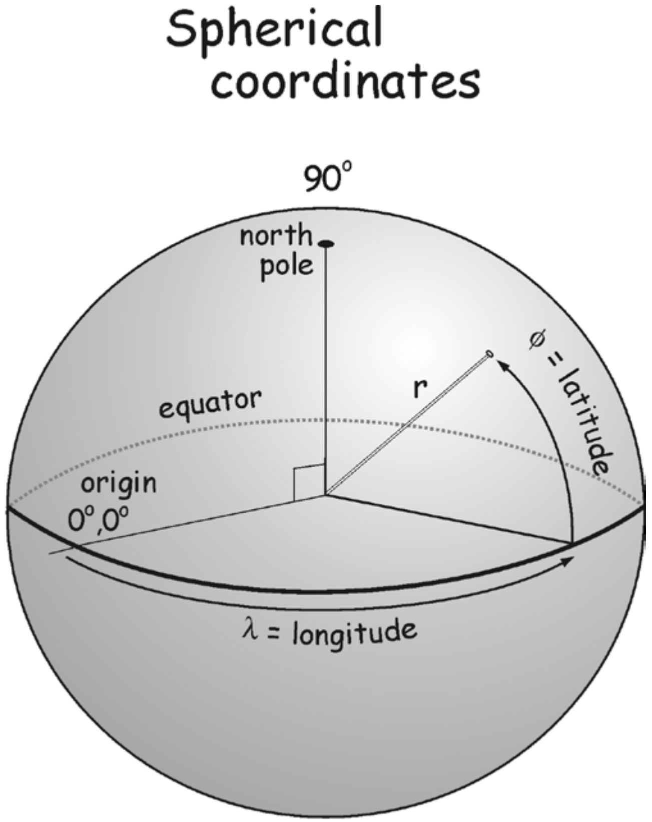
Bolstad 2012, Fig 2-3

2D and 3D

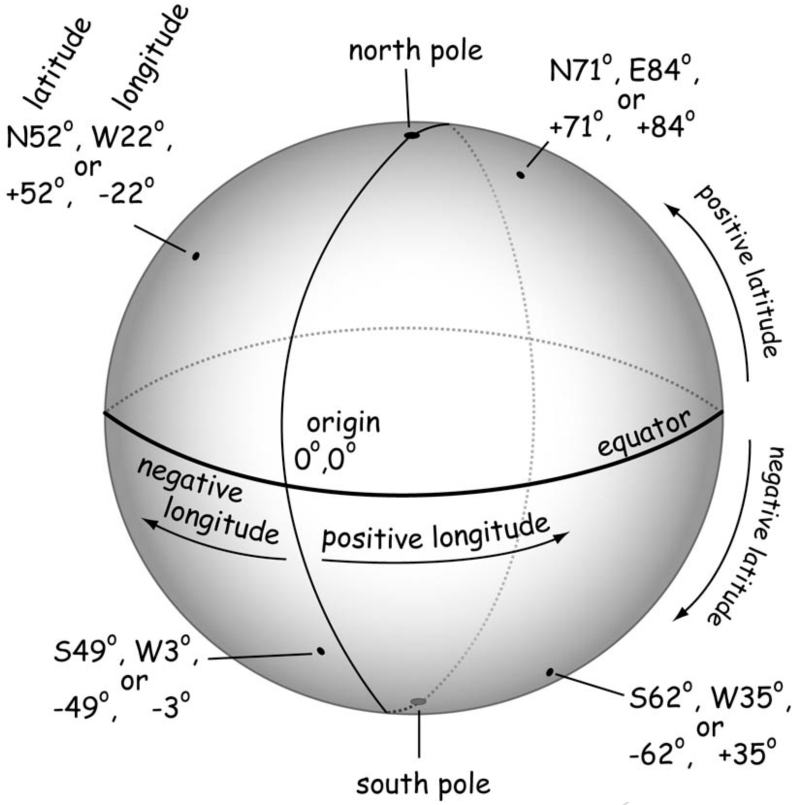


Bolstad 2012, Fig 2-5

Spherical Coordinate Systems

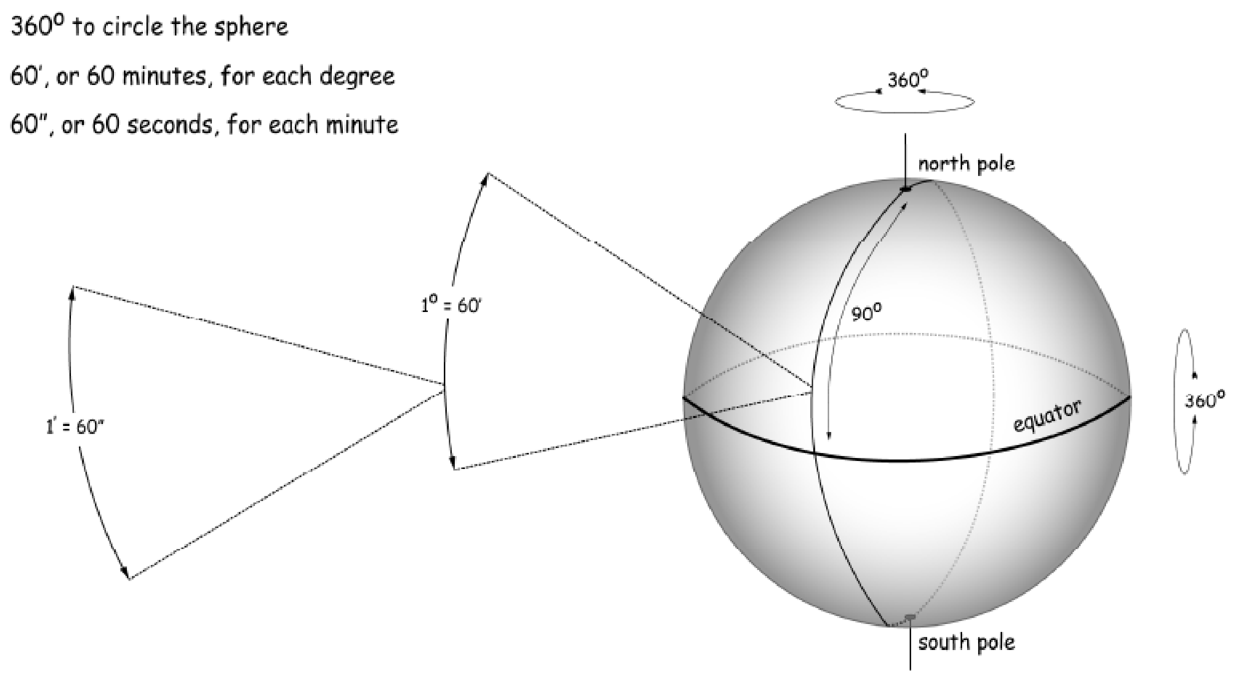


Bolstad 2012, Fig 2-6



Bolstad 2012, Fig 2-7

Degrees, Minutes, Seconds



Bolstad 2012, Fig 2-8

DD from DMS
DD = D + M/60 + S/3600
e.g.
DMS = 32° 45' 28"

DD = 32 + 45/60 + 28/3600
= 32 + 0.75 + 0.0077778
= 32.7577778

DMS from DD
D = integer part
M = integer of decimal part x 60
S = 2nd decimal x 60
e.g.
DD = 24.93547
D = 24
M = integer of 0.93547 x 60
= integer of 56.1282
= 56
S = 2nd decimal x 60
= 0.1282 * 60 = 7.692
so DMS is
24° 56' 7.692"

Bolstad 2012, Fig 2-9

Attribute Data Types

- Nominal – just named; no order (e.g. red, green, blue)
  - Binary large objects (BLOBs) – images and sound files
- Ordinal – ordered, but intervals are not equivalent (e.g. Large, Medium, Small; Likely, Indifferent, Unlikely)
  - Note: The difference in interest between someone who is “very likely” to do something and “likely” is not the same as the difference between “likely” and “indifferent”.
- Interval – numeric where intervals matter (e.g. temp. in degrees C or F; pH)
  - Note: 10° C is 5° hotter than 5° C, but it is not twice as hot.
- Ratio – numeric where 0 means “non” (e.g. height; weight; temp. in K)
  - Note: 10 K is twice as hot as 5 K

Domain – the set of possible values for a variable (e.g. red, green, blue; -10 to 10)

Vector vs Raster Data Models

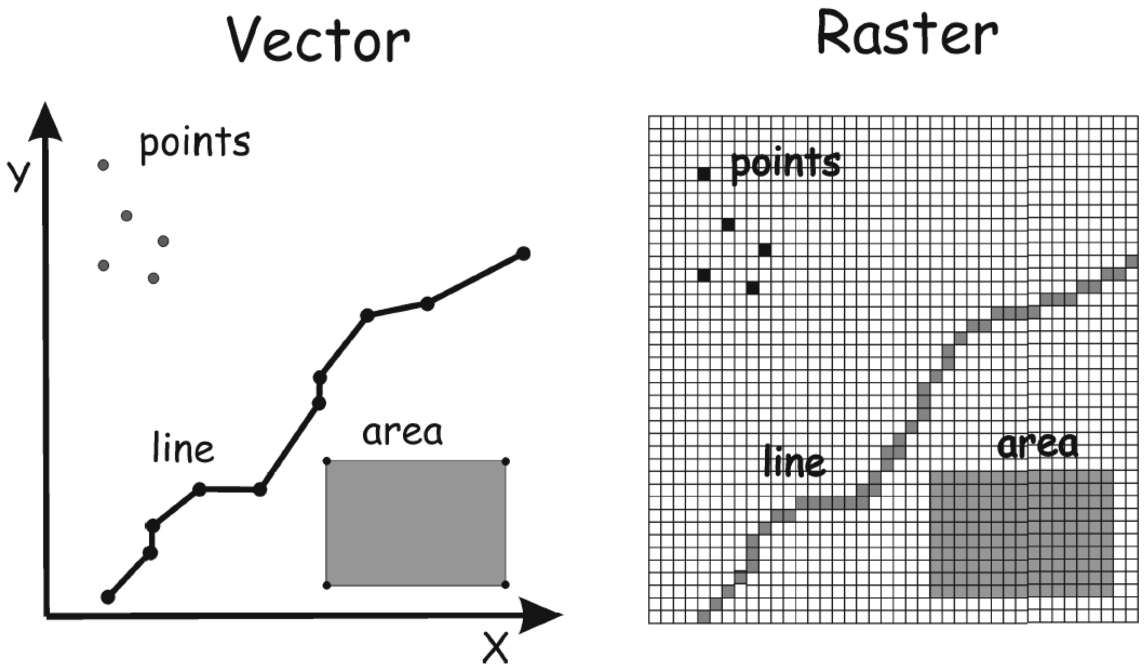
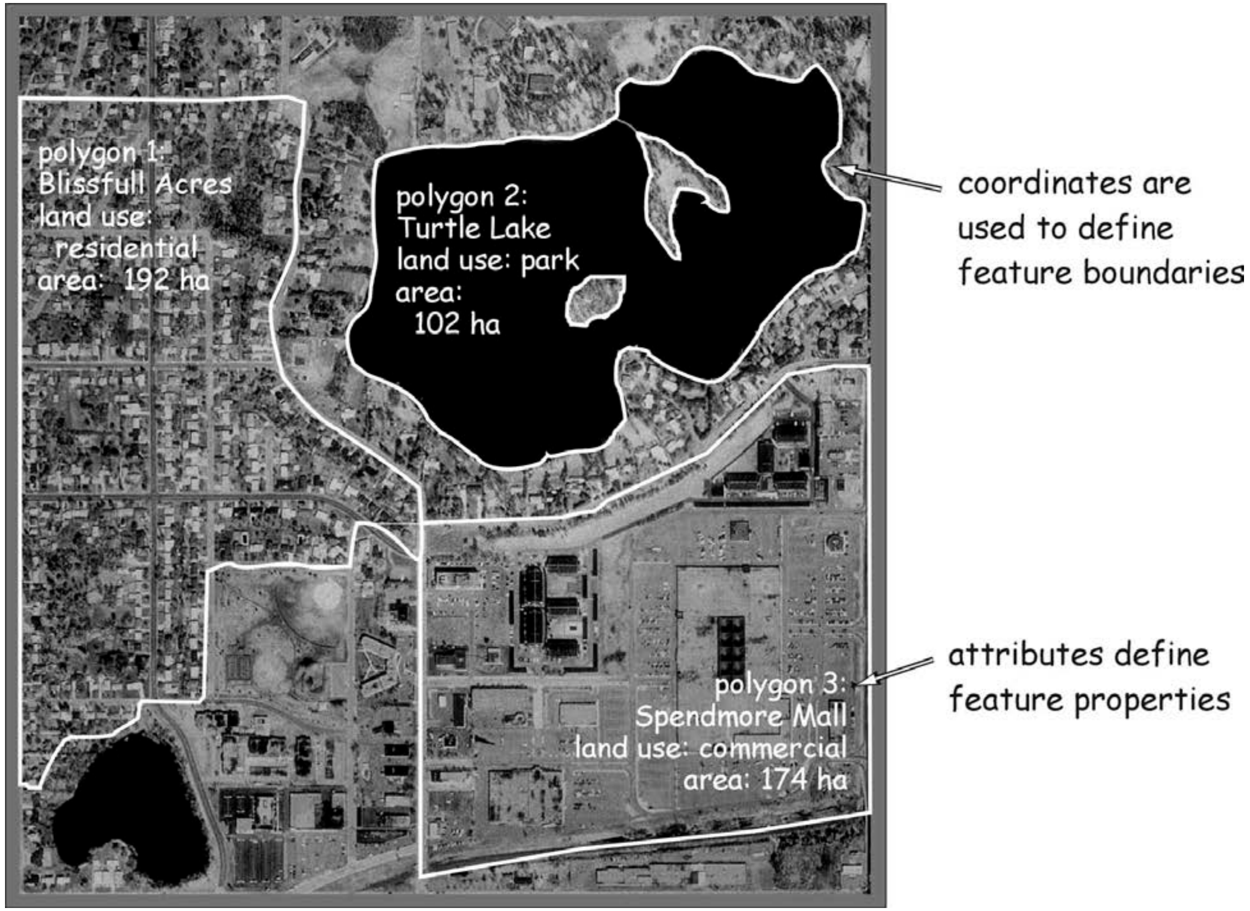


Figure 2-8: Vector and raster data models.

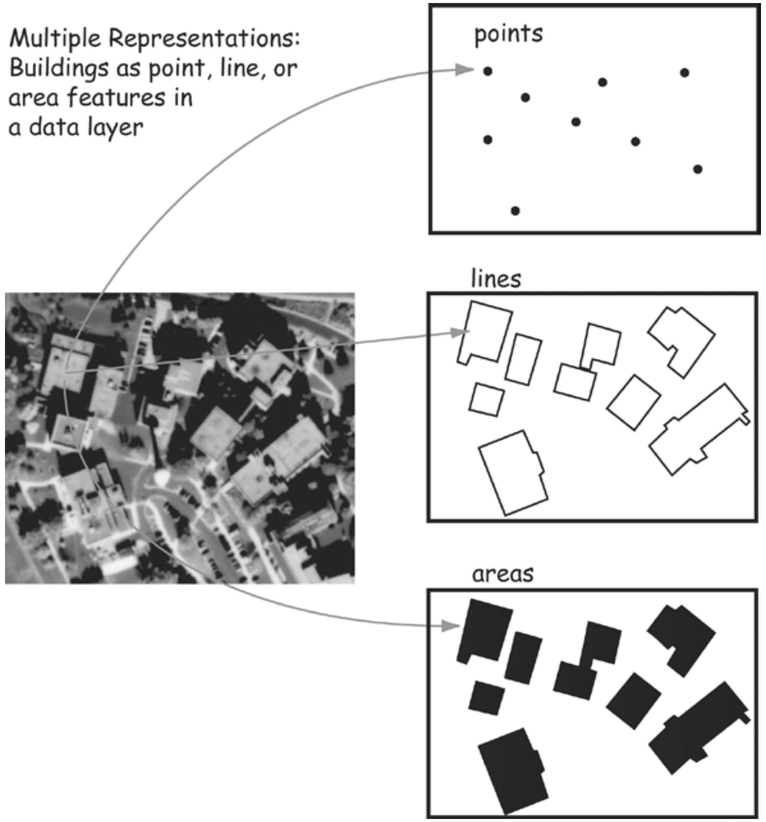
Bolstad 2012, Fig 2-11

Vectors on Top of a Raster



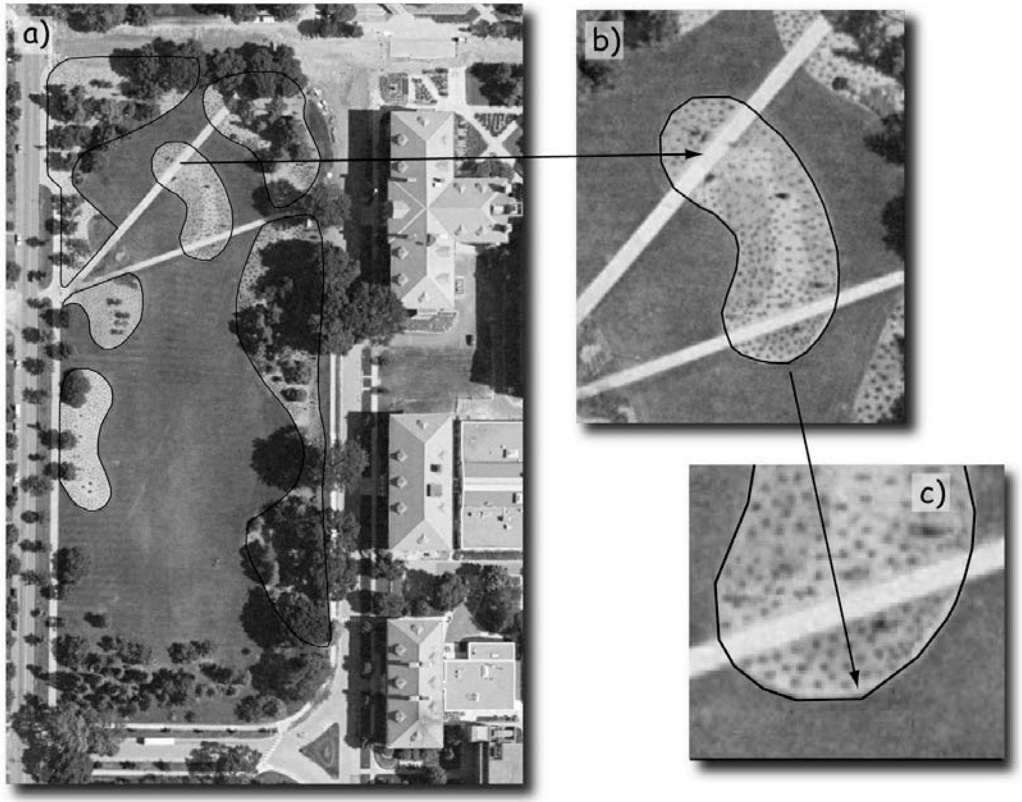
Bolstad 2012, Fig 2-12

Points, Lines, and Polygons



Bolstad 2012, Fig 2-14

Polygon Inclusion and Boundary Generalization

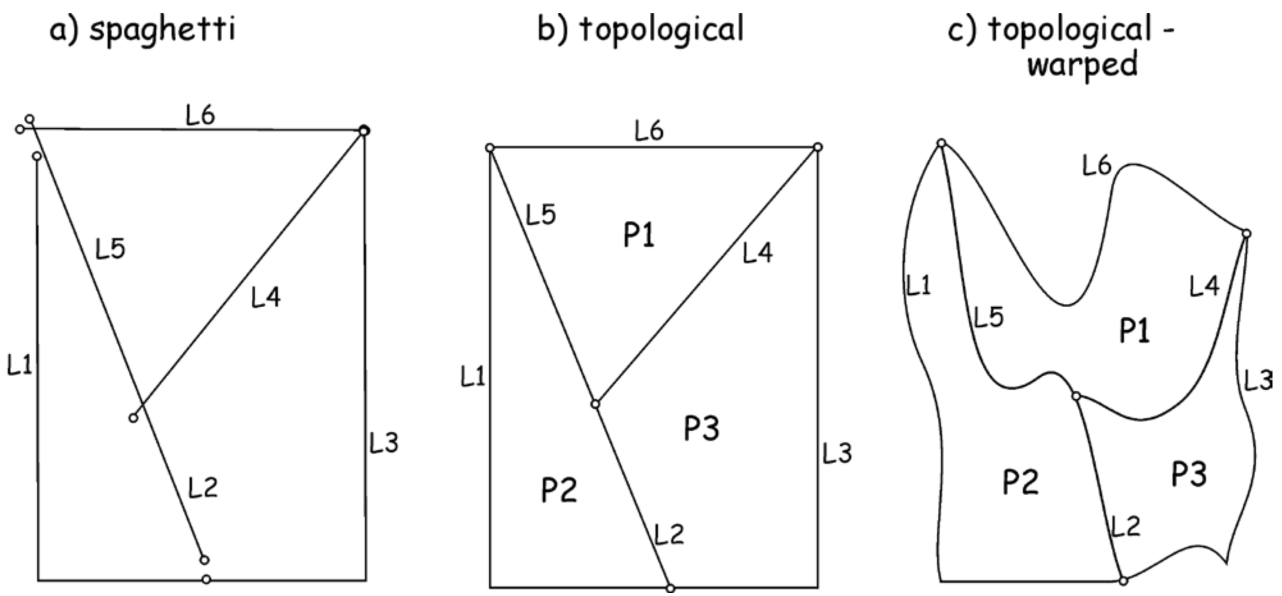


Bolstad 2012, Fig 2-15

Topology

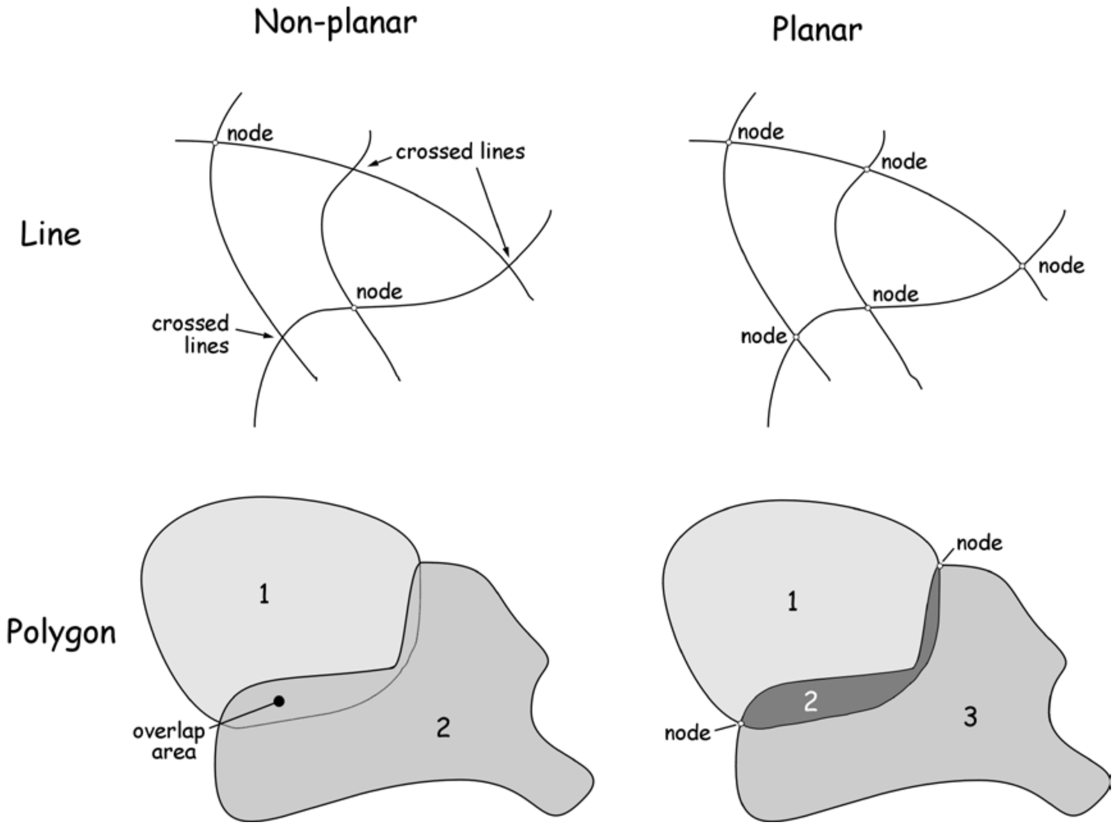
Topography – mapping of features that cover an area (think terrain and reliefs)

Topology – the way that features are related to one another in an area (think about the rules that define relationships, e.g. a county polygon needs to be completely and neatly contained in a state polygon)



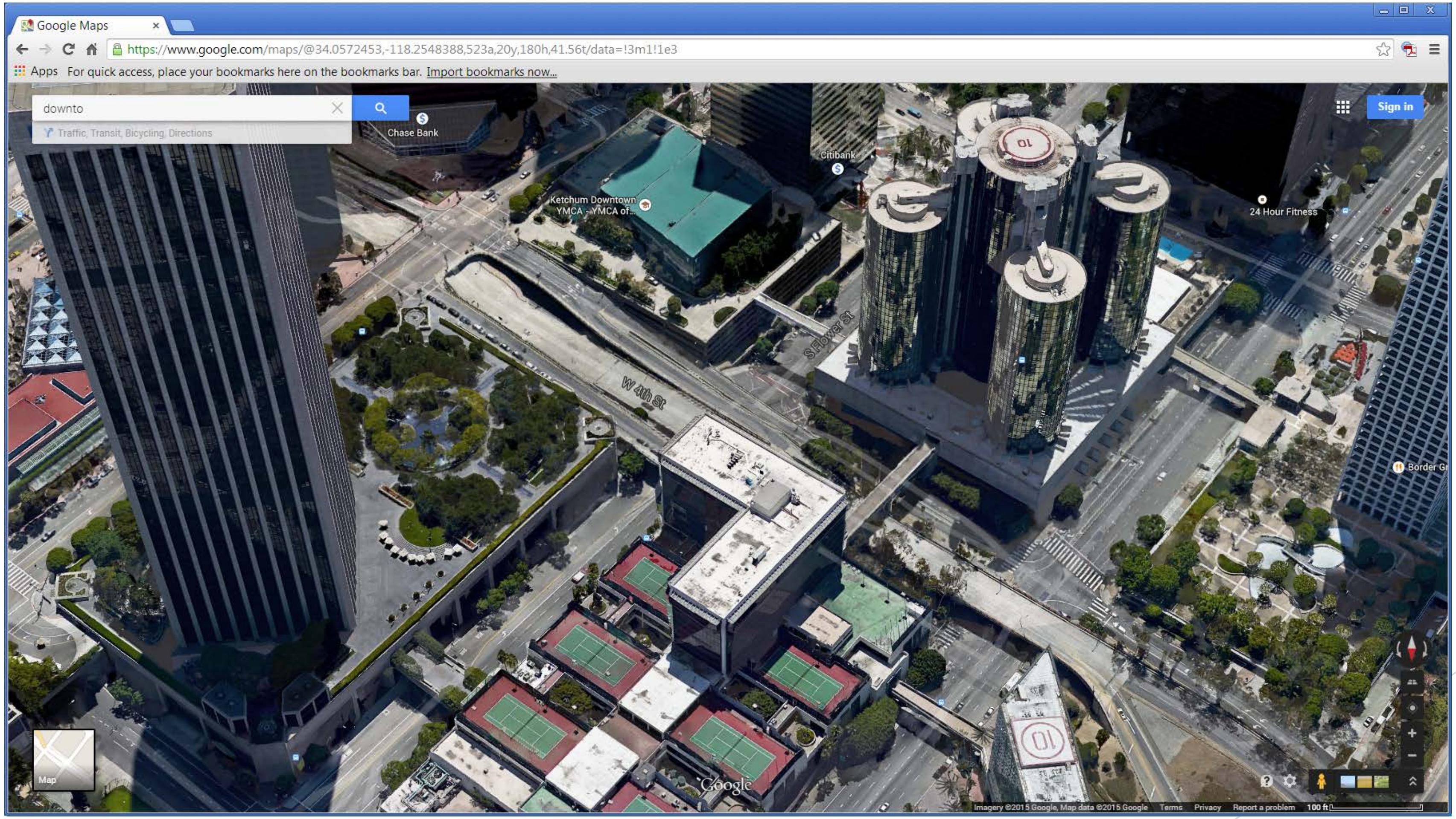
Bolstad 2012, Fig 2-16

More Topology

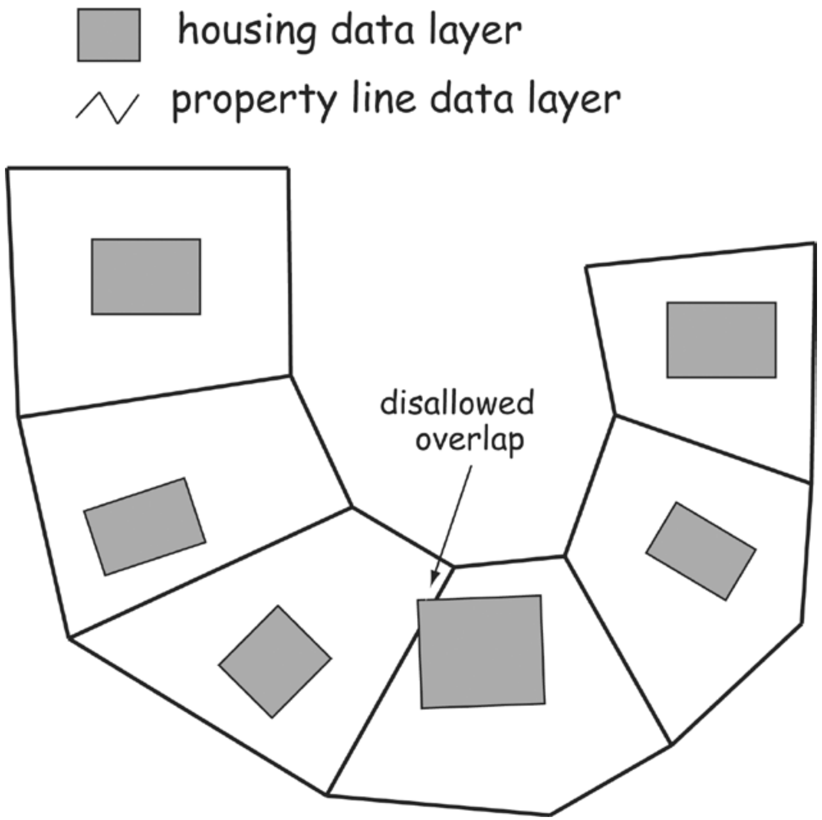


Bolstad 2012, Fig 2-17

Multiple Planes

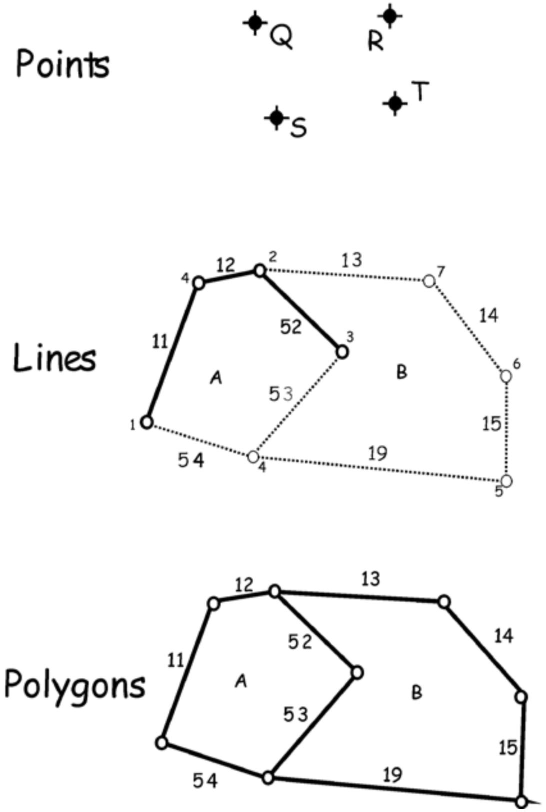


Topology Overlap Rules



Bolstad 2012, Fig 2-18

Simple Topology Examples



Point ID	X	Y
Q	32.7	45.6
R	76.3	19.5
S	22.7	15.8
etc...		

Line ID	Begin node	End node	Left poly	Right poly
11	1	4	...	A
12	4	2	...	A
52	2	3	B	A
etc ..				

Polygon ID	Lines
A	11,12,52,53,54
B	52,53,19, 15,14,13

Bolstad 2012, Fig 2-19

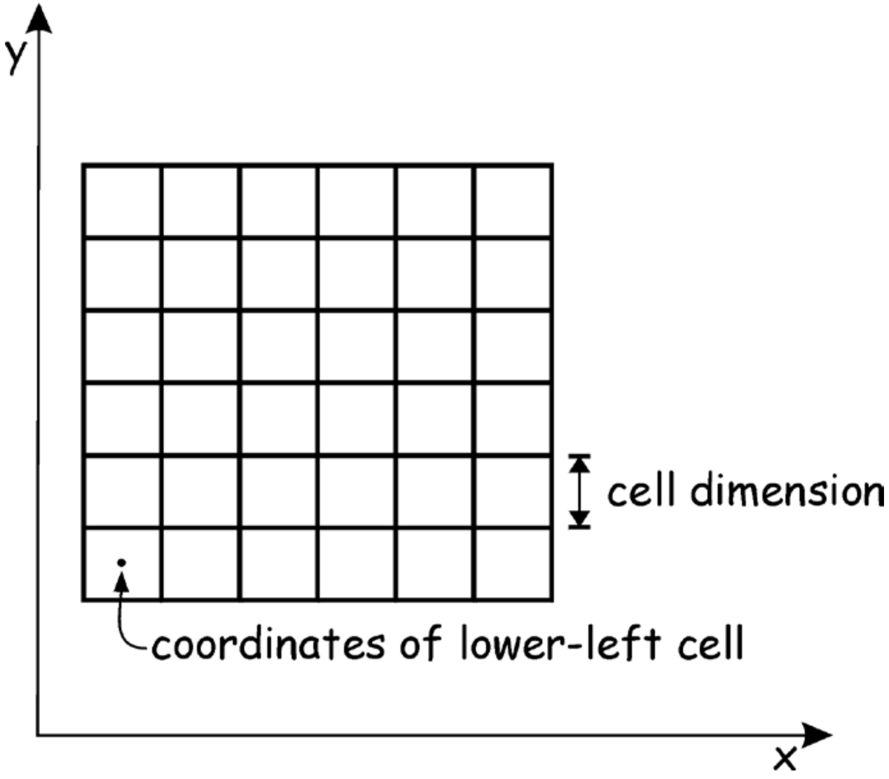
Raster Contents

a	a	a	a	r	f	f	a	a	a	a	a
a	a	a	a	r	f	f	a	a	a	a	a
a	a	a	f	r	f	f	a	a	a	a	a
a	a	a	r	r	f	f	a	a	a	a	a
a	a	a	r	f	f	f	a	a	a	a	a
a	f	f	r	f	f	f	a	a	a	a	a
a	f	f	r	f	u	f	a	a	a	a	a
h	h	h	h	h	h	h	h	h	h	h	h
f	f	r	u	u	u	u	a	a	a	a	a
f	f	r	f	u	u	a	a	a	a	a	a
f	f	f	r	f	f	a	a	a	a	a	a
f	f	f	f	r	f	a	a	a	a	a	a

a = agriculture
f = forest
h = highways
u = developed
r = river

Bolstad 2012, Fig 2-23

Important Defining Characteristics



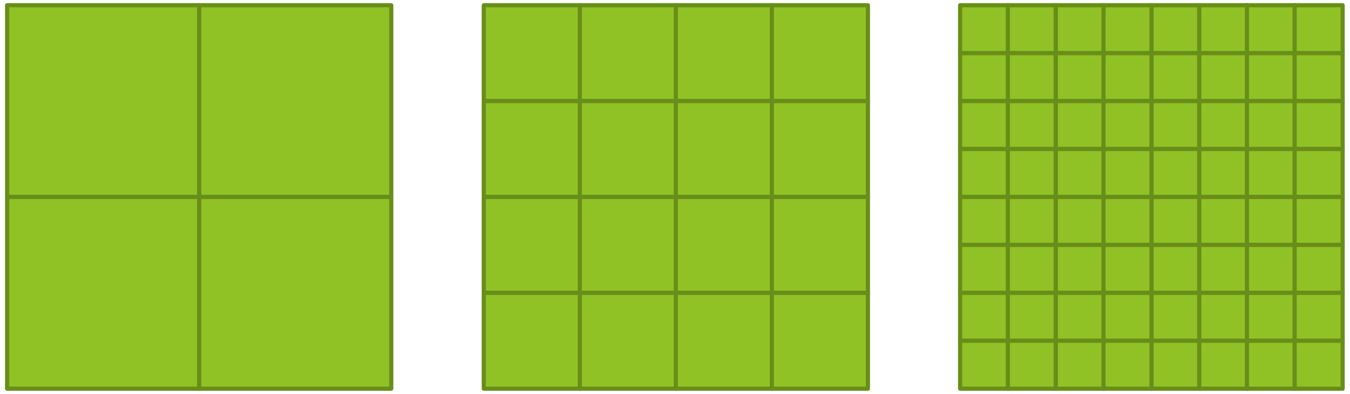
$$N_{cell} = N_{lower\text{-}left} + row * cell\ size \quad (2.2)$$

$$E_{cell} = E_{lower\text{-}left} + column * cell\ size \quad (2.3)$$

Bolstad 2012, Fig 2-21

Bolstad 2012, Eqs. 2.2 and 2.3

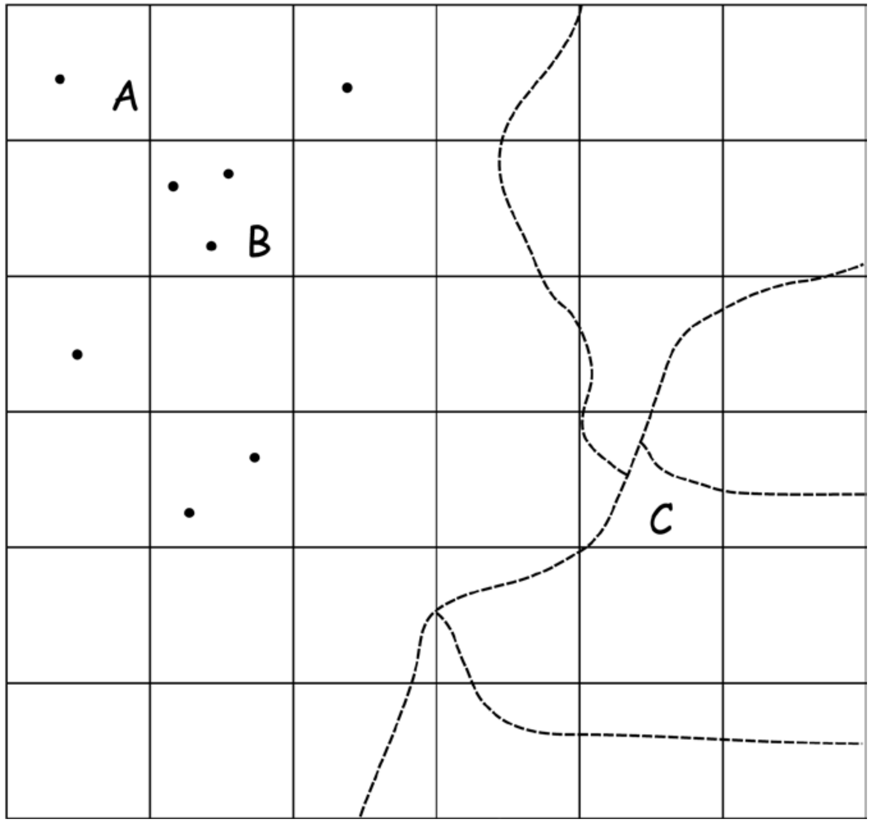
Raster Grids



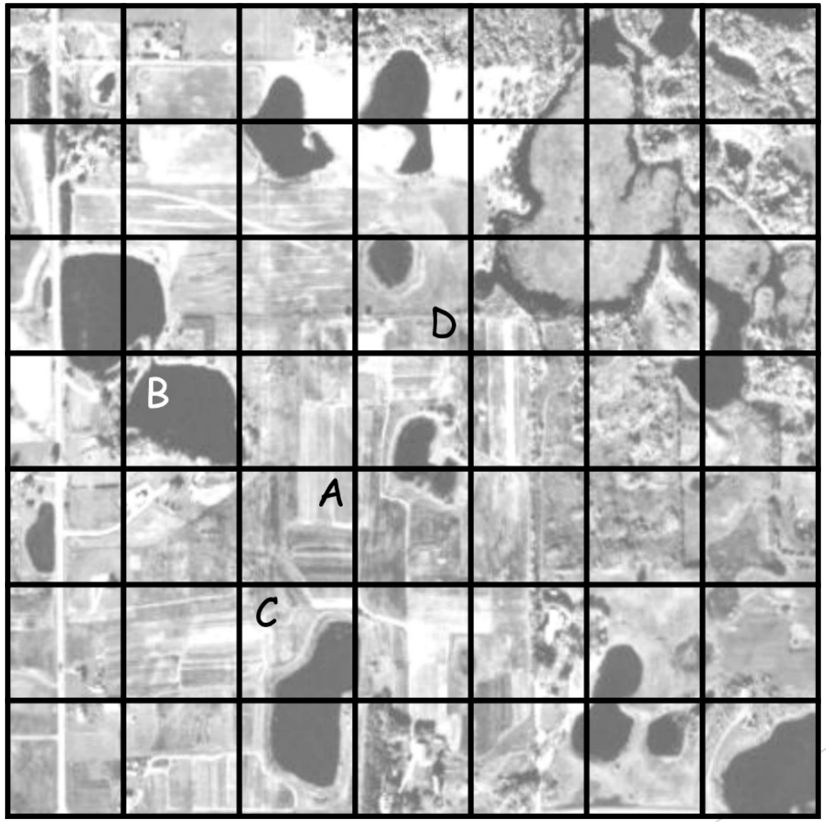
100 meter, 4 cells   50 meter, 16 cells   25 meter, 64 cells

Adapted from Bolstad 2012, Figure 2-22

Cell Assignment and Resolution



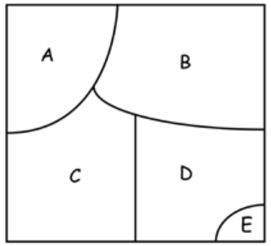
Bolstad 2012, Fig 2-24



Bolstad 2012, Fig 2-25

Raster and Vector Attribute Tables

a) Vector, one-to-one



attribute table

IDorg	class	area
A	10	16.8
B	11	22.2
C	15	18.4
D	21	16.4
E	10	3.8

b) Raster, one-to-one

A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	A	B	B	B	B	B	B
A	A	A	B	B	B	B	B	B	B
A	A	A	C	C	B	B	B	B	B
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	D	D
C	C	C	C	C	D	D	D	E	E
C	C	C	C	C	D	D	E	E	E

attribute table
(cell 1 is upper-left corner)

cell-ID	IDorg	class	area
1	A	10	0.8
2	A	10	0.8
3	A	10	0.8
4	A	10	0.8
5	B	11	0.8
6	B	11	0.8
7	B	11	0.8
⋮	⋮	⋮	⋮
100	E	10	0.8

c) Raster, many-to-one

10	10	10	10	11	11	11	11	11	11
10	10	10	10	11	11	11	11	11	11
10	10	10	10	11	11	11	11	11	11
10	10	10	11	11	11	11	11	11	11
10	10	10	15	15	11	11	11	11	11
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	21	21
15	15	15	15	15	21	21	21	10	10
15	15	15	15	15	21	21	10	10	10

attribute table

class	area
10	18.4
11	24.0
15	21.6
21	13.6

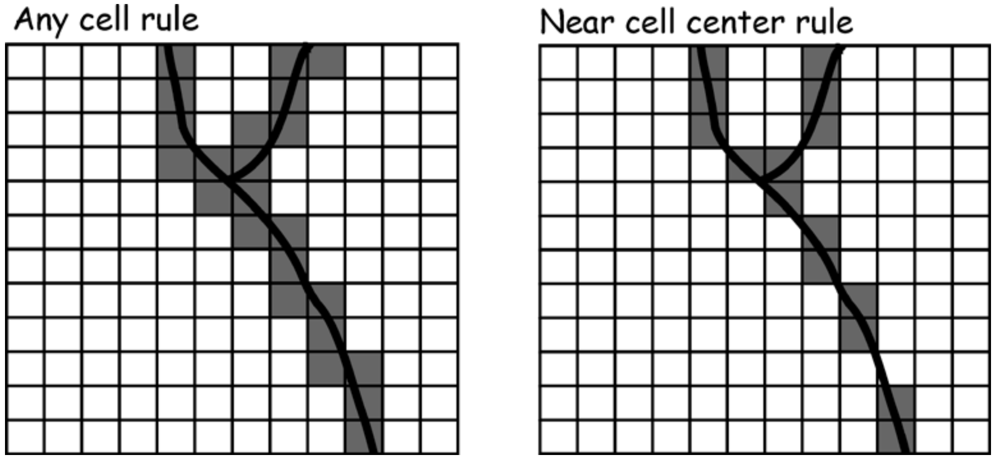
Bolstad 2012, Fig 2-26

More on Raster vs Vector

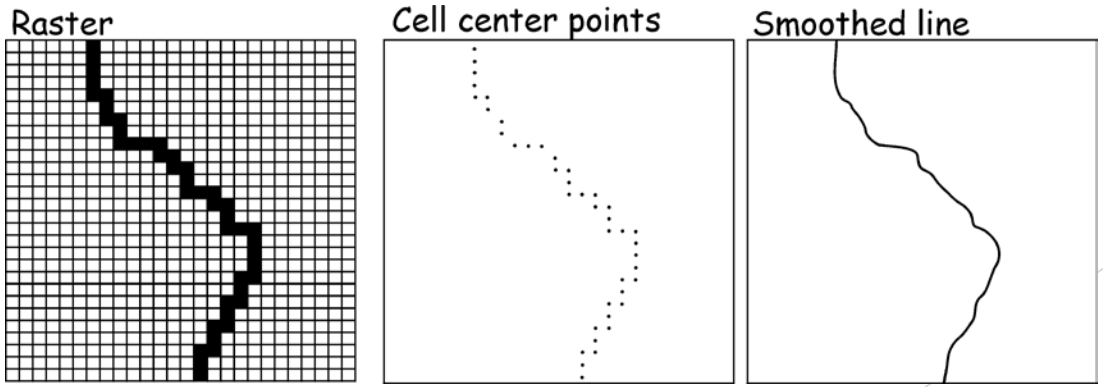
Characteristic	Raster	Vector
Data structure	Usually simple	Usually complex
Storage requirements	Larger for most datasets without compression	Smaller for most datasets
Coordinate conversion	May be slow due to data volumes, and require resampling	Simple
Analysis	Easy for continuous data, simple for many layer combinations	Preferred for network analyses, many other spatial operations more complex
Spatial precision	Floor set by cell size	Limited only by positional measurements
Accessibility	Easy to modify or program, due to simple data structure	Often complex
Display and output	Good for images, but discrete features may show “stairstep” edges	Map-like, with continuous curves, poor for images

Bolstad 2012, Table 2-2

Raster/Vector Conversions



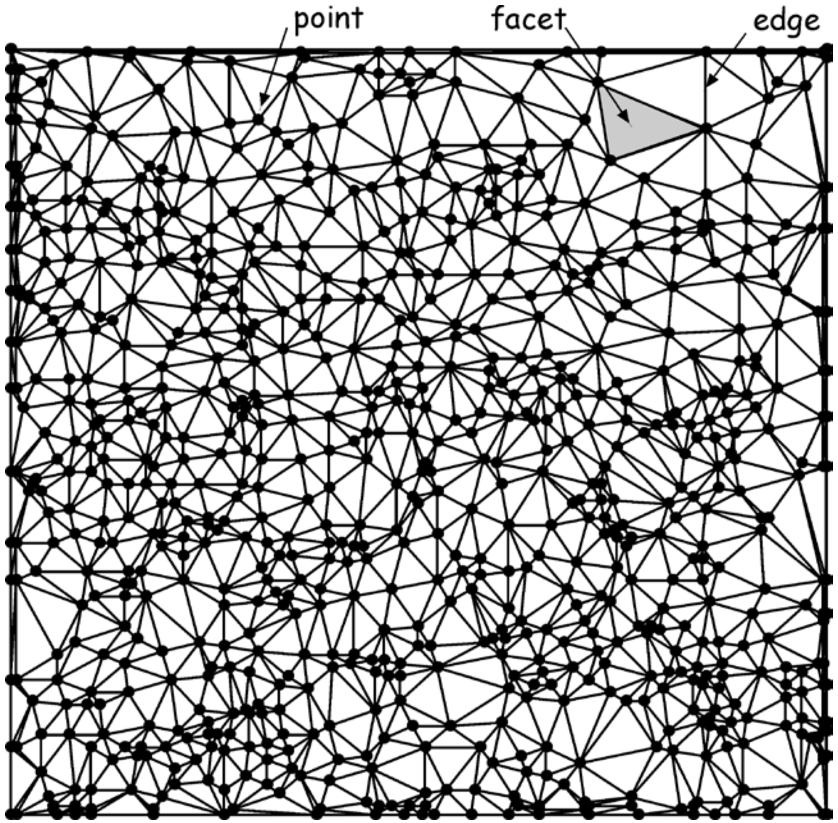
Two examples of vector to raster

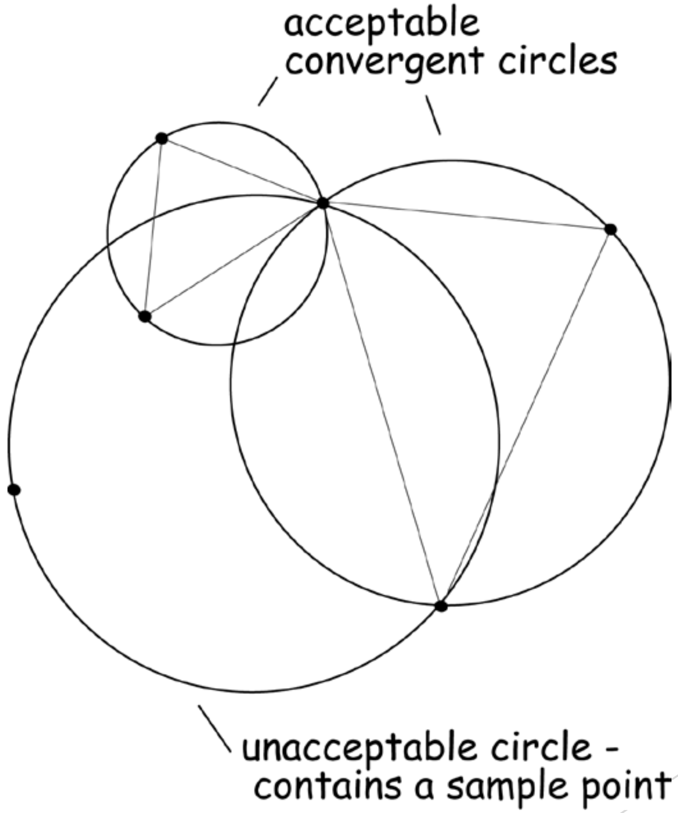


A process of raster to vector

Bolstad 2012, Fig 2-27

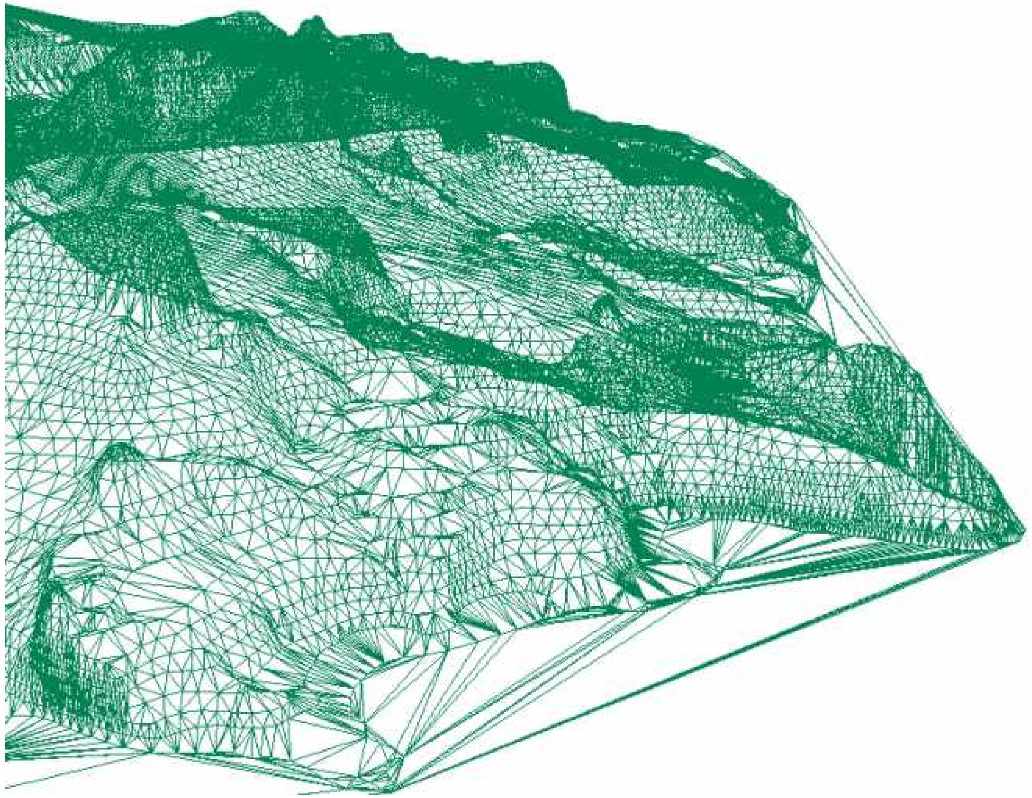
Triangulated Irregular Network (TIN) and Delaunay Triangles





Bolstad 2012, Fig 2-28

TINs for Elevation



http://www.geostat.iung.pulawy.pl

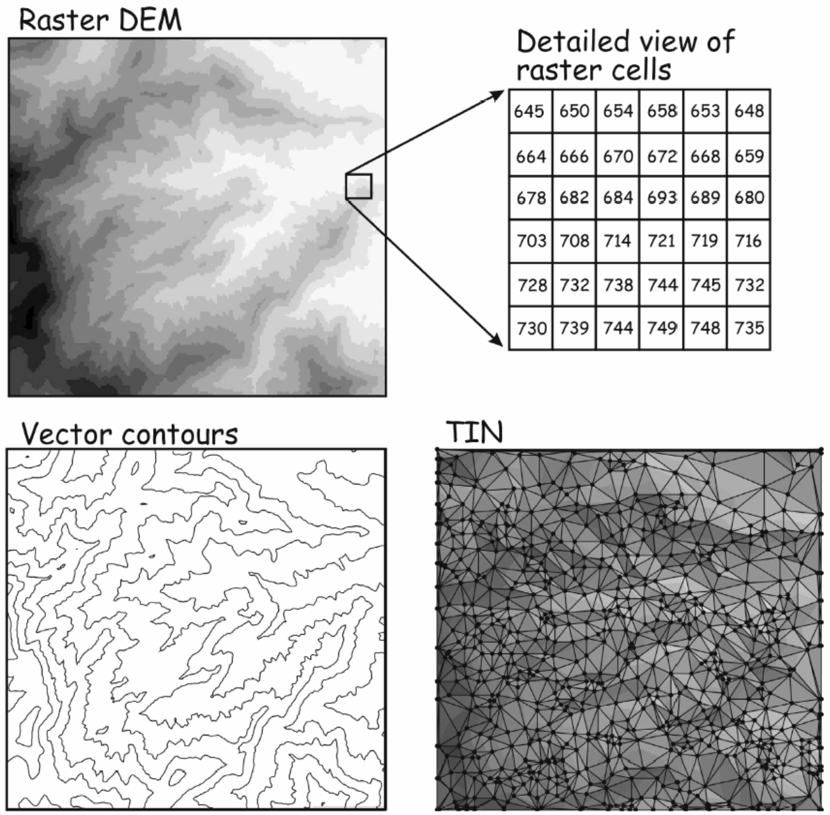
Different Data Models for the Same Entity

Elevation rasters and TINs are sometimes called:

DEM = Digital Elevation Model

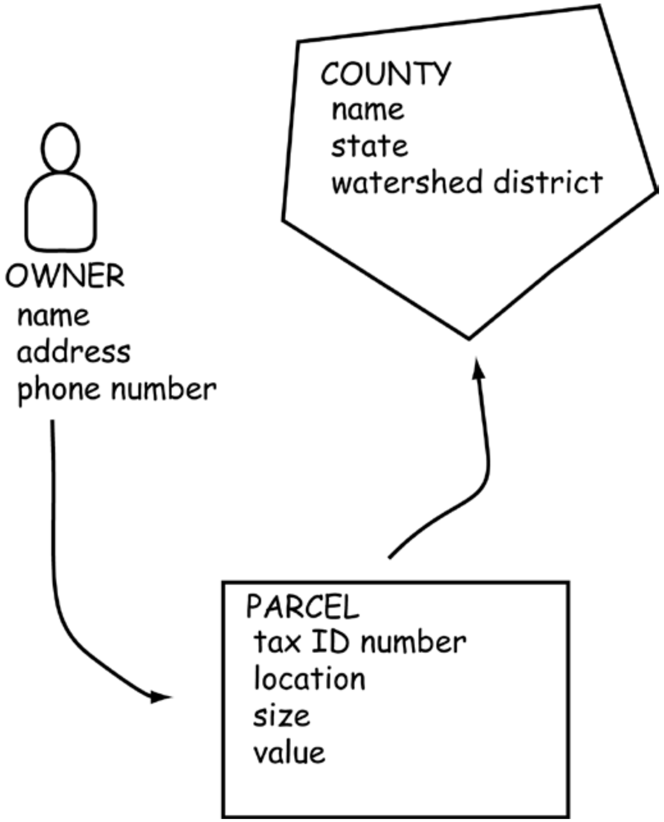
DTM = Digital Terrain Model

Note that you would not refer to vector contours as either a DEM or DTM

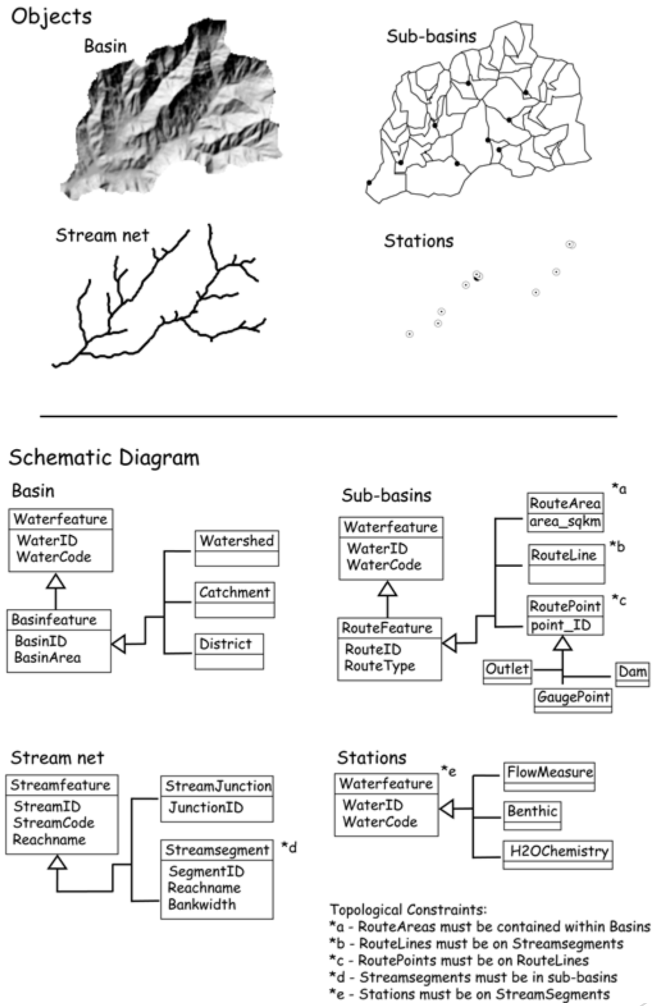


Bolstad 2012, Fig 2-29

Object Data Model



Bolstad 2012, Fig 2-30



Bolstad 2012, Fig 2-31

Data and File Structures

<u>Introduction to Binary</u>

Refresher on Exponents
$2^0 = 1$
$2^1 = 2$
$2^2 = 4$
$2^3 = 8$
$2^4 = ??$

Binary	Decimal
00000001	1
00000010	2
00000011	3
00000100	4
00000101	5
00000111	6
00001000	7
00001001	8
0000????	9

So 1101 in binary translates to:

$(1 \times 2^3) + (1 \times 2^2) + (0 \times 2^1) + (1 \times 2^0) =$
$8 + 4 + 0 + 1 = 13$

So what does 1111 in binary translate to?

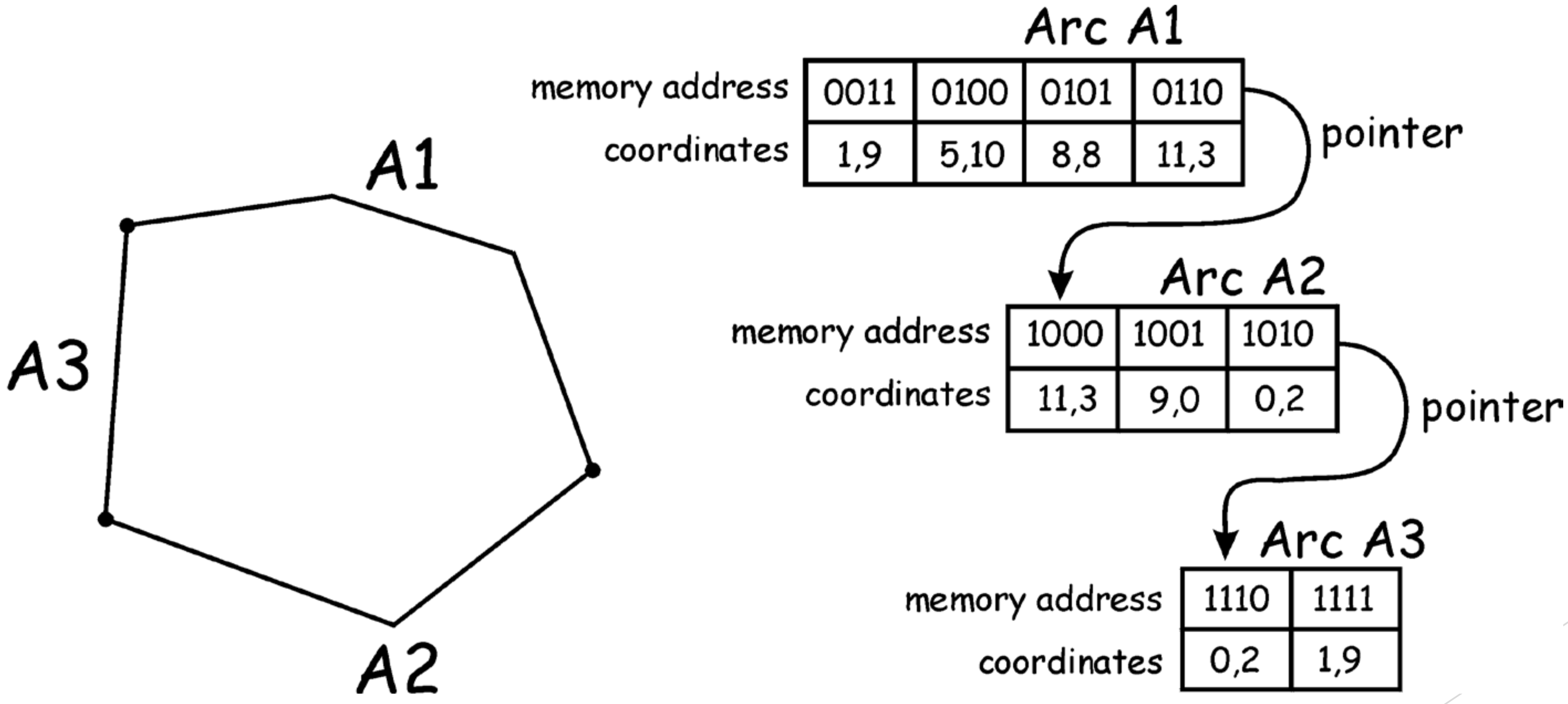
bit = 1 column in binary
byte = 8 bits = 8 columns in binary, stores values from 0-255

ASCII and ANSI

- American Standard Code for Information Interchange (ASCII) is 7 bits = 127 total values
- American National Standards Institute (ANSI) is 8 bits = 255 values
- Each value represents a character, such as “a”

- The point: you may see data files that are indicated as ASCII or ANSI, and now you know what these are.

Pointers and Indexes



Bolstad 2012, Fig 2-33

Shapefiles

- Mandatory files
  - .shp — shape format; the feature geometry itself
  - .shx — shape index format; a positional index of the feature geometry to allow seeking forwards and backwards quickly
  - .dbf — attribute format; columnar attributes for each shape, in dBase IV format
- Other files
  - .prj — projection format; the coordinate system and projection information, a plain text file describing the projection using well-known text format
  - .sbn and .sbx — a spatial index of the features
  - .fbn and .fbx — a spatial index of the features that are read-only
  - .ain and .aih — an attribute index of the active fields in a table
  - .ixs — a geocoding index for read-write datasets
  - .mxs — a geocoding index for read-write datasets (ODB format)
  - .atx — an attribute index for the .dbf file in the form of shapefile.columnname.atx (ArcGIS 8 and later)
  - .shp.xml — geospatial metadata in XML format, such as ISO 19115 or other XML schema
  - .cpg — used to specify the code page (only for .dbf) for identifying the character encoding to be used
  - .qix — an alternative quadtree spatial index used by MapServer and GDAL/OGR software

List taken from http://en.wikipedia.org/wiki/Shapefile on 01/20/2015

Data Compression

Run Length Codes

Raster

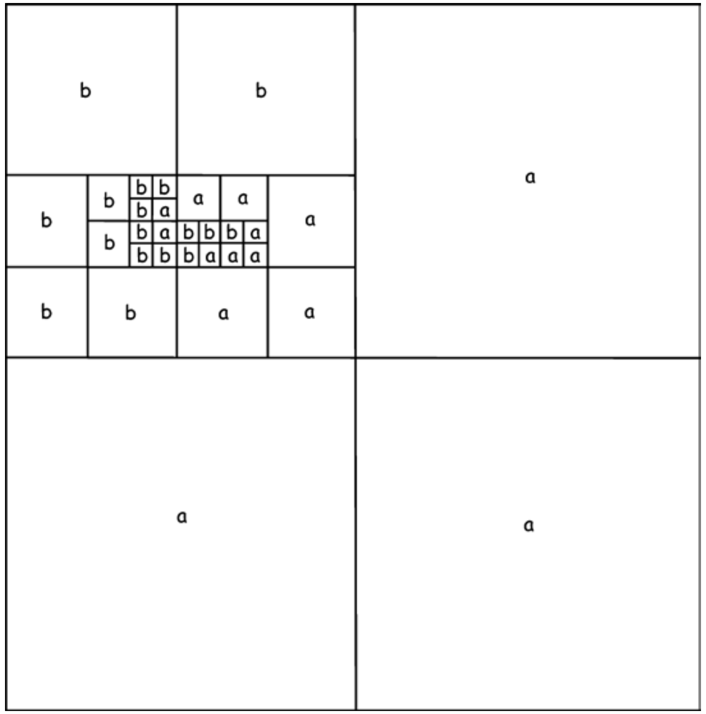
9	9	6	6	6	6	6	7
6	6	6	6	6	6	6	6
9	9	6	6	6	6	7	7
9	8	9	6	6	7	7	5

Run-length codes

2:9, 5:6, 1:7
8:6
2:9, 4:6, 2:7
1:9, 1:8, 1:9, 2:6, 2:7, 1:5

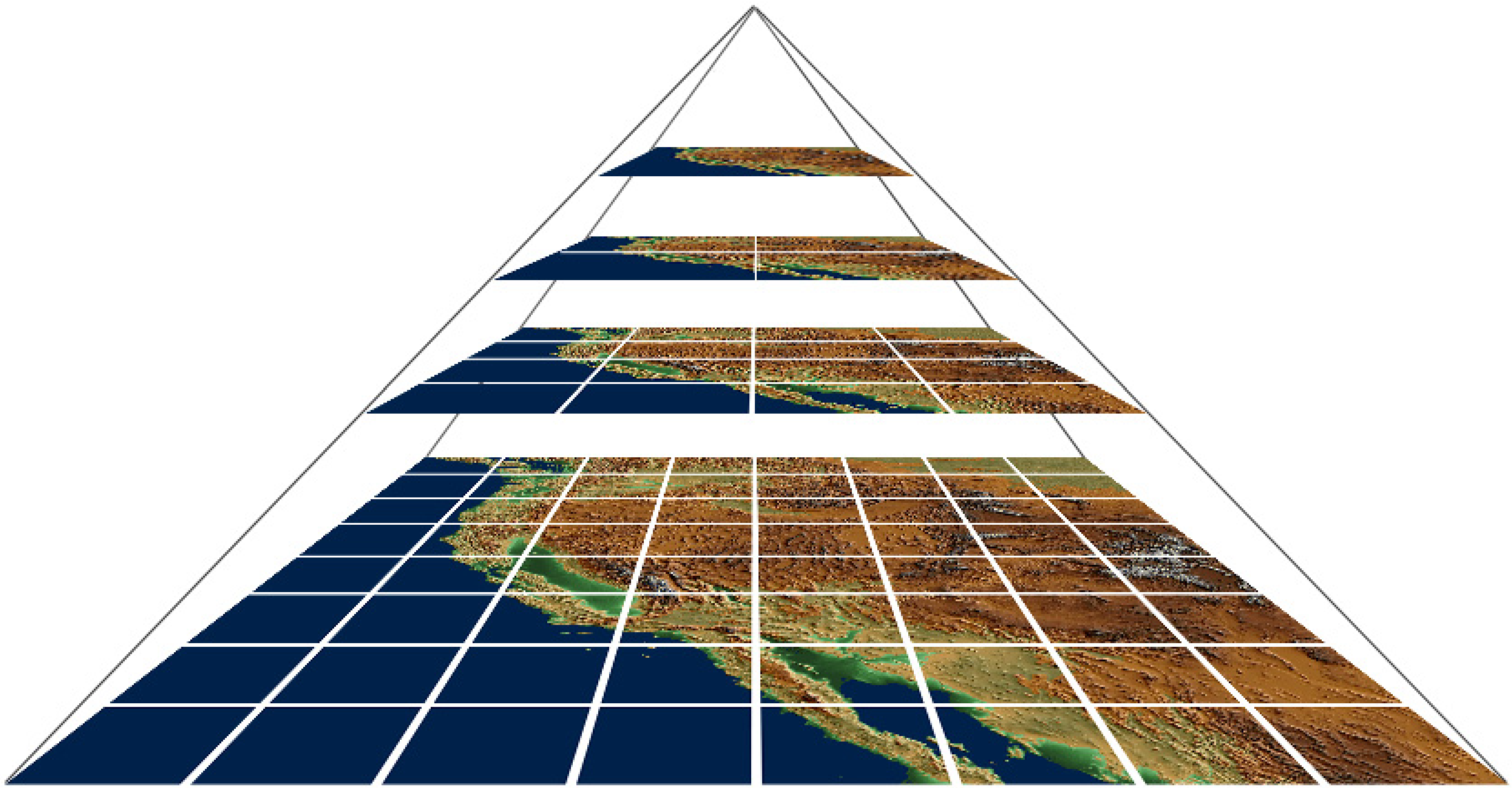
Bolstad 2012, Fig 2-34

Quadtrees



Bolstad 2012, Fig 2-35

Raster Pyramids



http://wiki.osgeo.org/wiki/GDAL2Tiles_SoC_2007

Lesson Roadmap

2a) List the data requirements → 2b) Examine data → 2c) Reframe the problem statement

1) Frame problem and explore study area	2) Preview data	3) Choose the data
4) Build the database	5) Edit data	6) Conduct the analysis
7) Automate the analysis	8) Prepare analysis results	9) Share results online