GIS data models

Agenda

- Geographic data
- Vector data
- Raster data

Components of a GIS

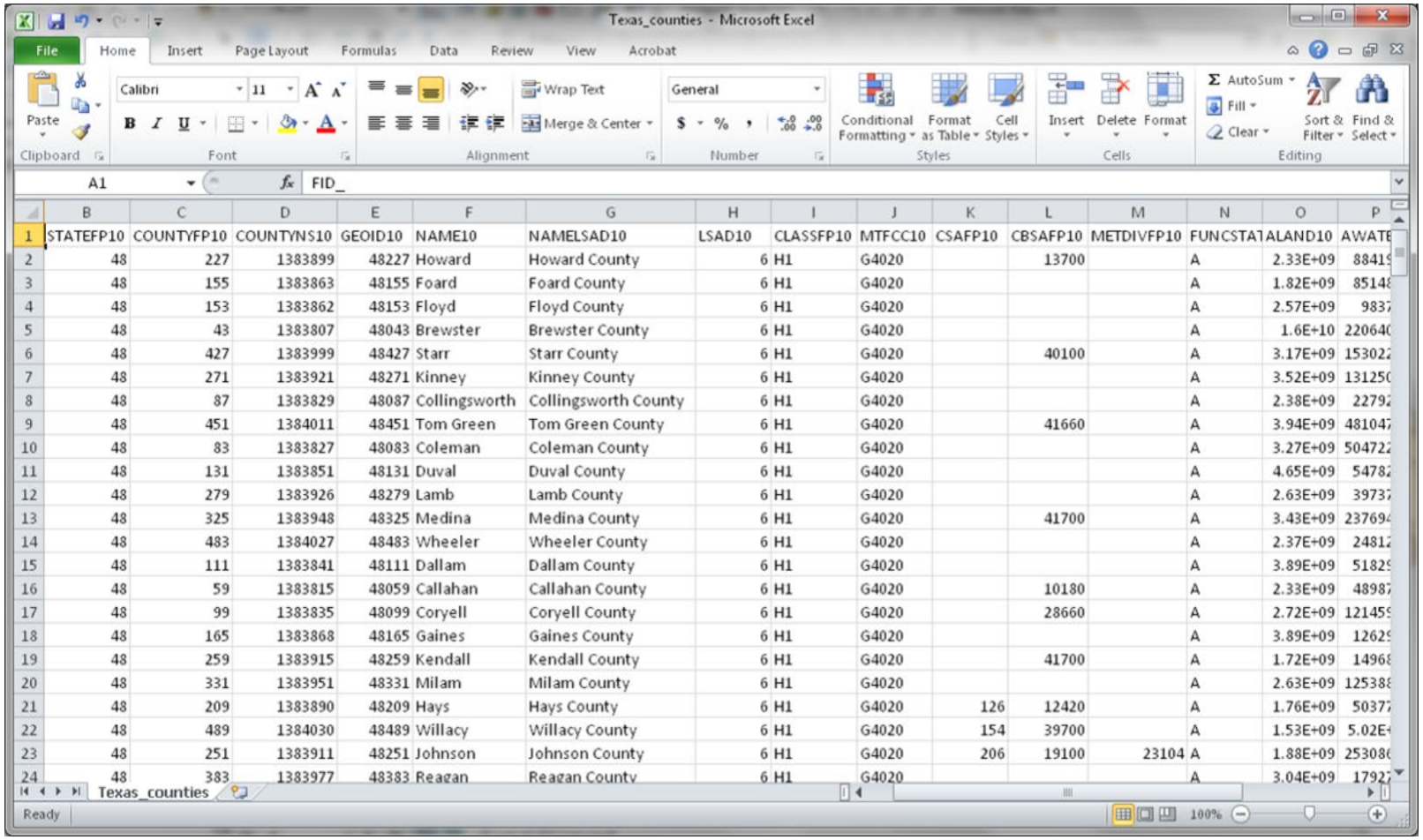
- Hardware
- Software
- Data
- Users

GIS software

- Desktop GIS: ArcGIS Desktop, ArcGIS Pro, QGIS, MapInfo
- Web GIS: ArcGIS Online, MapBox, CartoDB, Other
- applications:
- Database (e.g. PostgreSQL/PostGIS), statistical software (e.g. R), programming languages (e.g. Python)

Data

- The **information** component of GIS
- Example: Texas county data in Excel spreadsheet



	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	STATEFP10	COUNTYFP10	COUNTYNS10	GEOID10	NAME10	NAMELSAD10	LSAD10	CLASSFP10	MTFCC10	CSAFP10	CBSAFP10	METDIVFP10	FUNCSTAT	ALAND10	AWATE
2	48	227	1383899	48227	Howard	Howard County	6	H1	G4020		13700		A	2.33E+09	8841
3	48	155	1383863	48155	Foard	Foard County	6	H1	G4020				A	1.82E+09	8514
4	48	153	1383862	48153	Floyd	Floyd County	6	H1	G4020				A	2.57E+09	983
5	48	43	1383807	48043	Brewster	Brewster County	6	H1	G4020				A	1.6E+10	22064
6	48	427	1383999	48427	Starr	Starr County	6	H1	G4020		40100		A	3.17E+09	15302
7	48	271	1383921	48271	Kinney	Kinney County	6	H1	G4020				A	3.52E+09	13125
8	48	87	1383829	48087	Collingsworth	Collingsworth County	6	H1	G4020				A	2.38E+09	2279
9	48	451	1384011	48451	Tom Green	Tom Green County	6	H1	G4020		41660		A	3.94E+09	48104
10	48	83	1383827	48083	Coleman	Coleman County	6	H1	G4020				A	3.27E+09	50472
11	48	131	1383851	48131	Duval	Duval County	6	H1	G4020				A	4.65E+09	5478
12	48	279	1383926	48279	Lamb	Lamb County	6	H1	G4020				A	2.63E+09	3973
13	48	325	1383948	48325	Medina	Medina County	6	H1	G4020		41700		A	3.43E+09	23769
14	48	483	1384027	48483	Wheeler	Wheeler County	6	H1	G4020				A	2.37E+09	2481
15	48	111	1383841	48111	Dallam	Dallam County	6	H1	G4020				A	3.89E+09	5182
16	48	59	1383815	48059	Callahan	Callahan County	6	H1	G4020		10180		A	2.33E+09	4898
17	48	99	1383835	48099	Coryell	Coryell County	6	H1	G4020		28660		A	2.72E+09	12145
18	48	165	1383868	48165	Gaines	Gaines County	6	H1	G4020				A	3.89E+09	1262
19	48	259	1383915	48259	Kendall	Kendall County	6	H1	G4020		41700		A	1.72E+09	1496
20	48	331	1383951	48331	Milam	Milam County	6	H1	G4020				A	2.63E+09	12538
21	48	209	1383890	48209	Hays	Hays County	6	H1	G4020	126	12420		A	1.76E+09	5037
22	48	489	1384030	48489	Willacy	Willacy County	6	H1	G4020	154	39700		A	1.53E+09	5.02E+
23	48	251	1383911	48251	Johnson	Johnson County	6	H1	G4020	206	19100	23104	A	1.88E+09	25308
24	48	383	1383977	48383	Reagan	Reagan County	6	H1	G4020				A	3.04E+09	1792

Geographic data

- Geographic data includes **reference to position on Earth's surface**
- Example: Tarrant County

Geographic data

"FID" and "Shape" fields store information about geographic position…

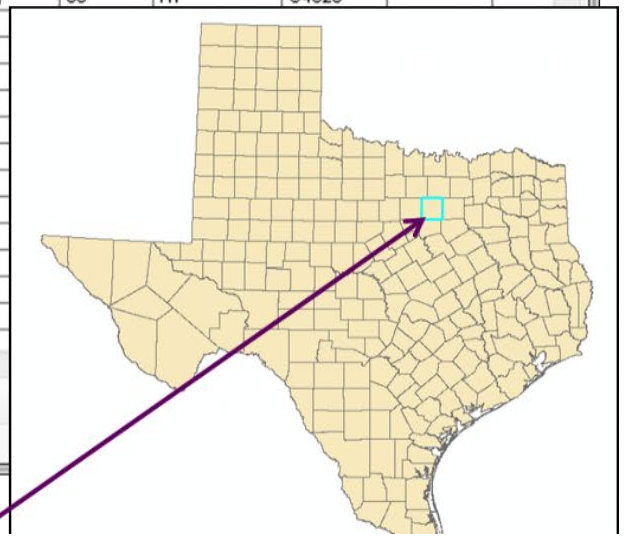
Table

Texas_counties

FID	Shape	STATEFP10	COUNTYFP10	COUNTYNS10	GEOID10	NAME10	NAMELSAD10	LSAD10	CLASSFP10	MTFCC10	CSAFP10	CBSA
24	Polygon	48	417	01383994	48417	Shackelford	Shackelford County	06	H1	G4020		
249	Polygon	48	419	01383995	48419	Shelby	Shelby County	06	H1	G4020		
180	Polygon	48	421	01383996	48421	Sherman	Sherman County	06	H1	G4020		
210	Polygon	48	423	01383997	48423	Smith	Smith County	06	H1	G4020	540	46340
27	Polygon	48	425	01383998	48425	Somervell	Somervell County	06	H1	G4020	206	24180
4	Polygon	48	427	01383999	48427	Starr	Starr County	06	H1	G4020		40100
32	Polygon	48	429	01384000	48429	Stephens	Stephens County	06	H1	G4020		
45	Polygon	48	431	01384001	48431	Sterling	Sterling County	06	H1	G4020		
138	Polygon	48	433	01384002	48433	Stonewall	Stonewall County	06	H1	G4020		
120	Polygon	48	435	01384003	48435	Sutton	Sutton County	06	H1	G4020		
48	Polygon	48	437	01384004	48437	Swisher	Swisher County	06	H1	G4020		
93	Polygon	48	439	01384005	48439	Tarrant	Tarrant County	06	H1	G4020	206	19100
29	Polygon	48	441	01384006	48441	Taylor	Taylor County	06	H1	G4020		10180
67	Polygon	48	443	01384007	48443	Terrell	Terrell County	06	H1	G4020		
101	Polygon	48	445	01384008	48445	Terry	Terry County	06	H1	G4020		
62	Polygon	48	447	01384009	48447	Throckmorton	Throckmorton County	06	H1	G4020		
220	Polygon	48	449	01384010	48449	Titus	Titus County					
7	Polygon	48	451	01384011	48451	Tom Green	Tom Green County					
57	Polygon	48	453	01384012	48453	Travis	Travis County					
191	Polygon	48	455	01384013	48455	Trinity	Trinity County					
186	Polygon	48	457	01384014	48457	Tyler	Tyler County					
229	Polygon	48	459	01384015	48459	Upshur	Upshur County					
84	Polygon	48	461	01384016	48461	Upton	Upton County					
61	Polygon	48	463	01384017	48463	Uvalde	Uvalde County					
35	Polygon	48	465	01384018	48465	Val Verde	Val Verde County					
217	Polygon	48	467	01384019	48467	Van Zandt	Van Zandt County					
72	Polygon	48	469	01384020	48469	Victoria	Victoria County					
185	Polygon	48	471	01384021	48471	Walker	Walker County					
197	Polygon	48	473	01384022	48473	Waller	Waller County					

0 (1 out of 254 Selected)

Texas_counties



which can then be visualized on a map

GIS data displayed as layers

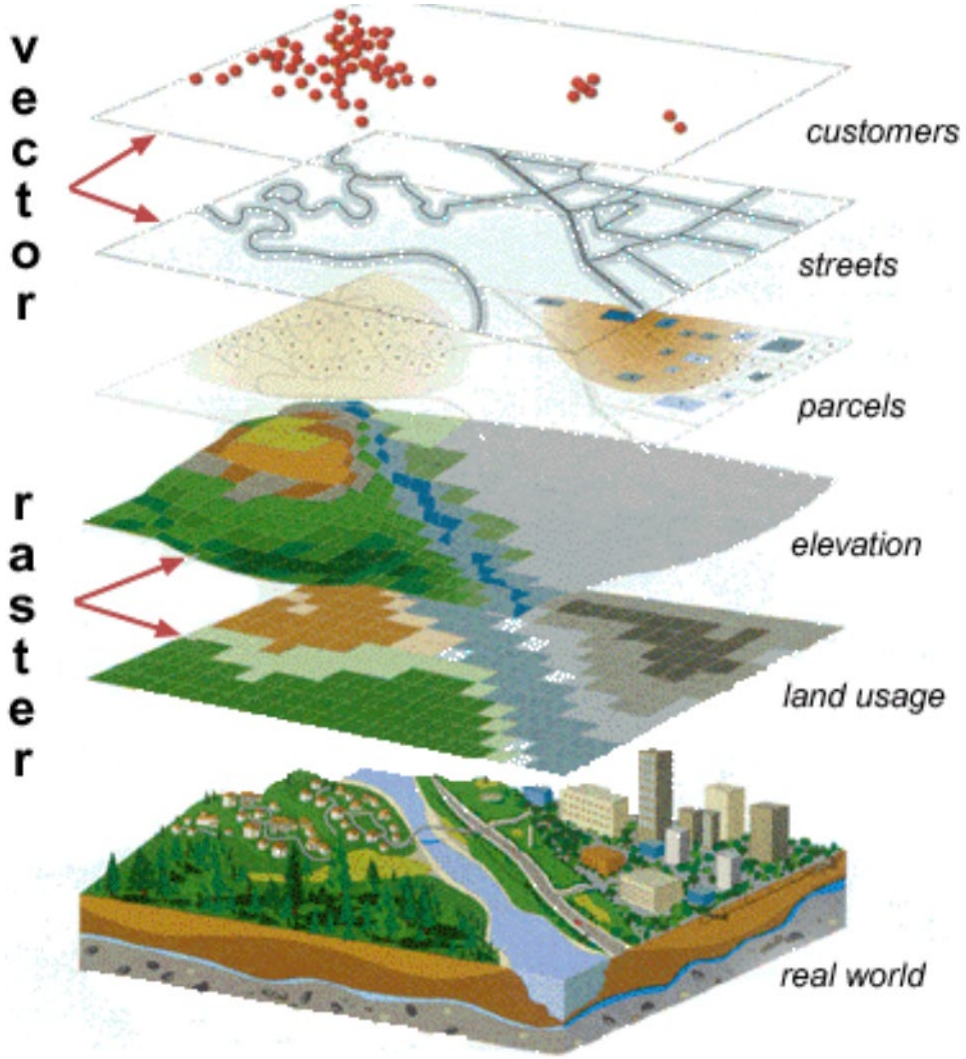
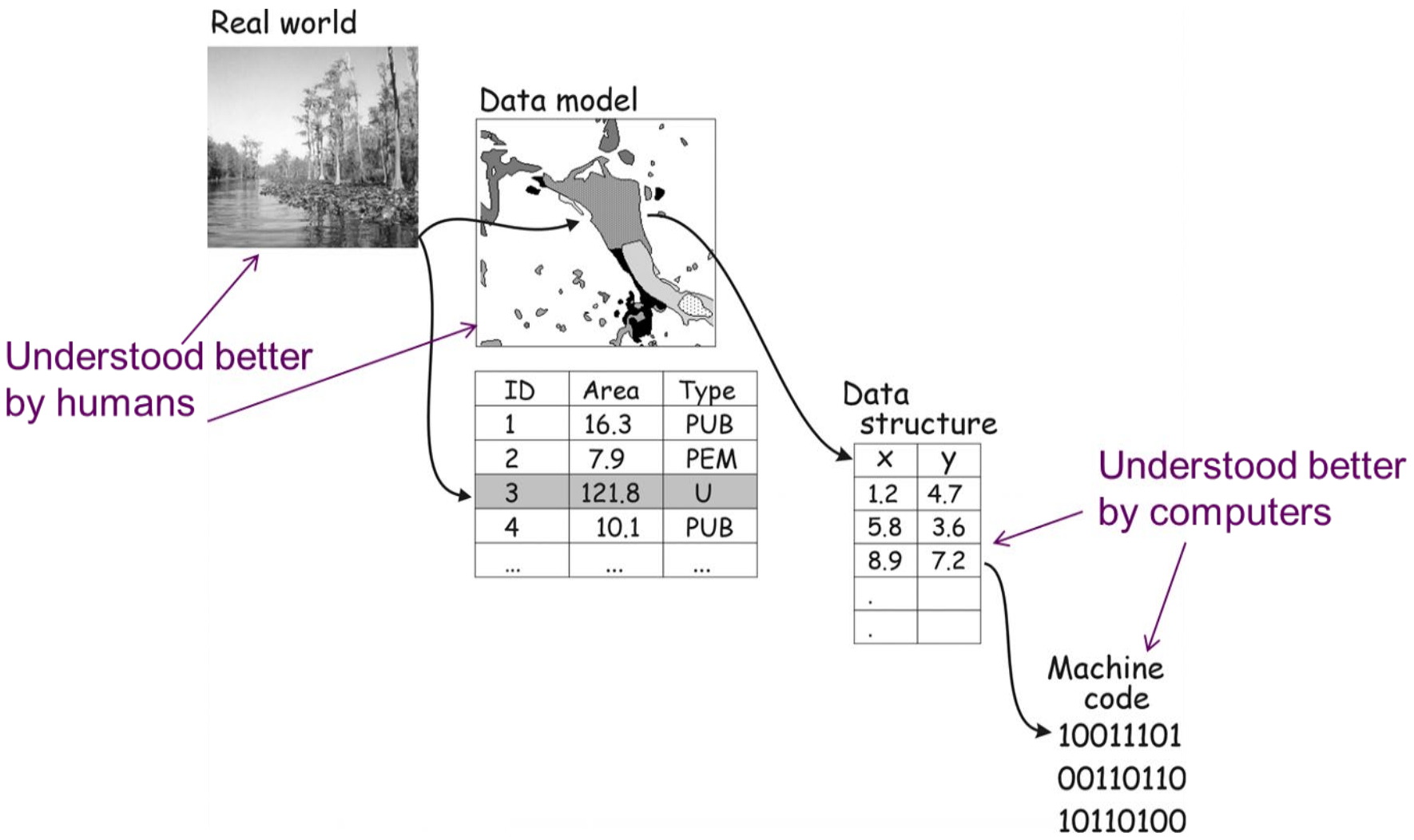


Image source: Wexford County, MI

Data models: abstraction of reality

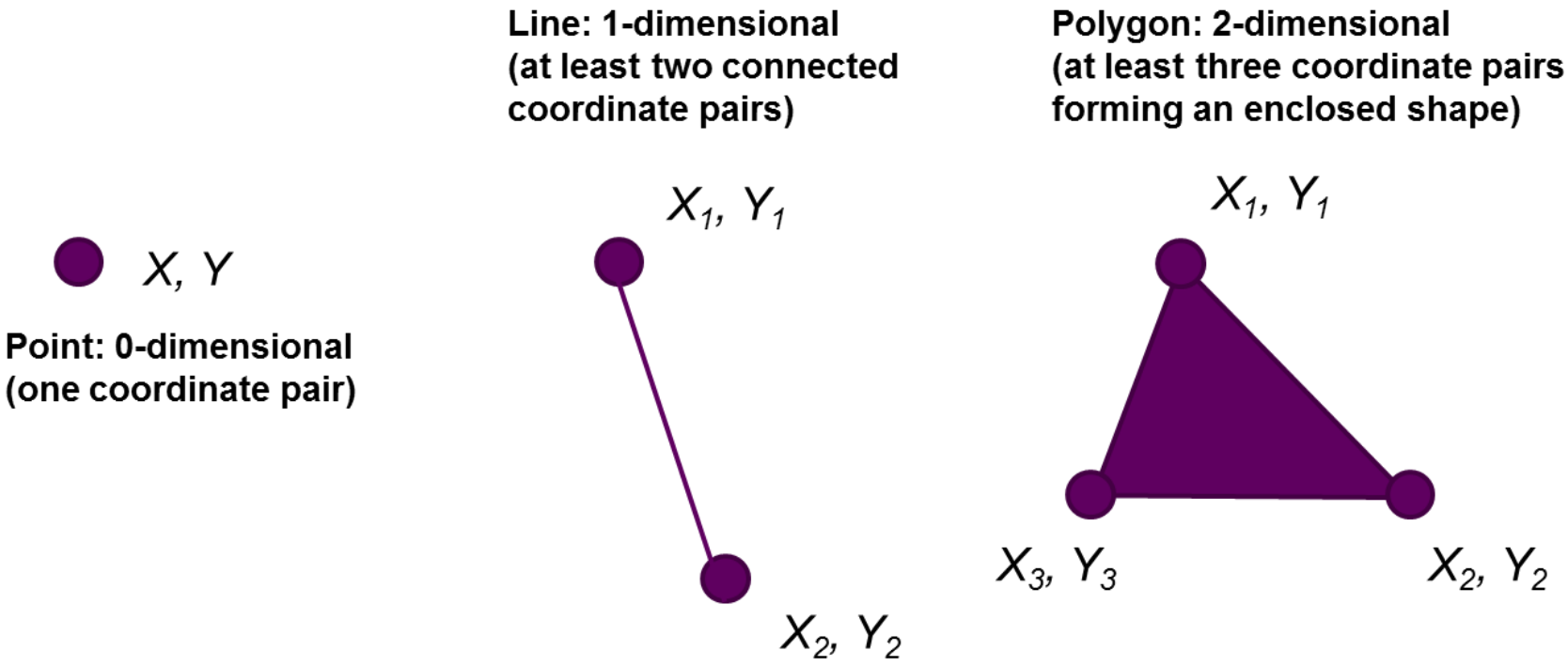


Types of GIS data

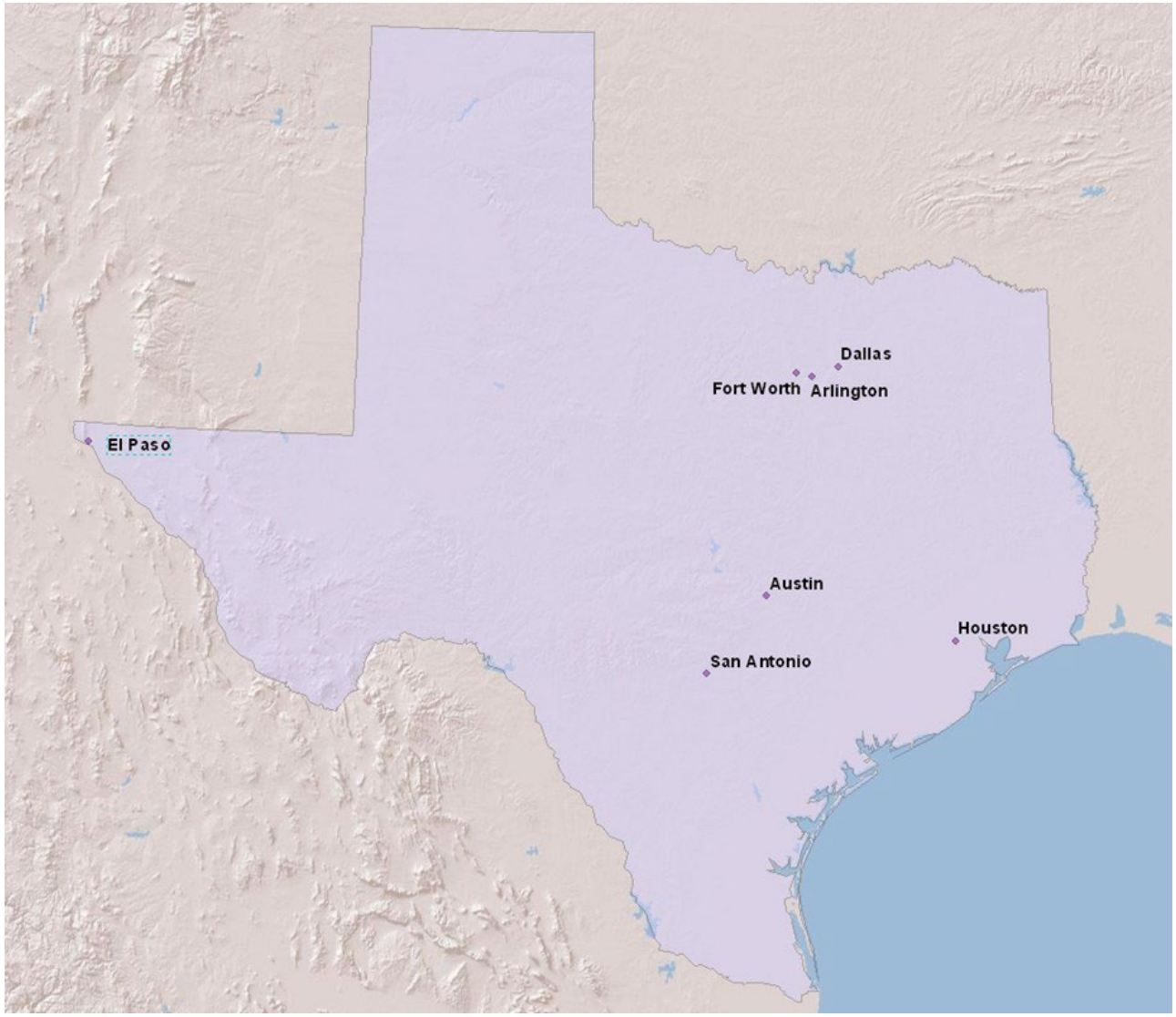
- Vector data: used to represent discrete geographic features
  - Examples: cities, lakes, roads, parcels, buildings
- Raster data: used to represent continuous geographic phenomena
  - Examples: imagery, elevation, temperature

Vector data

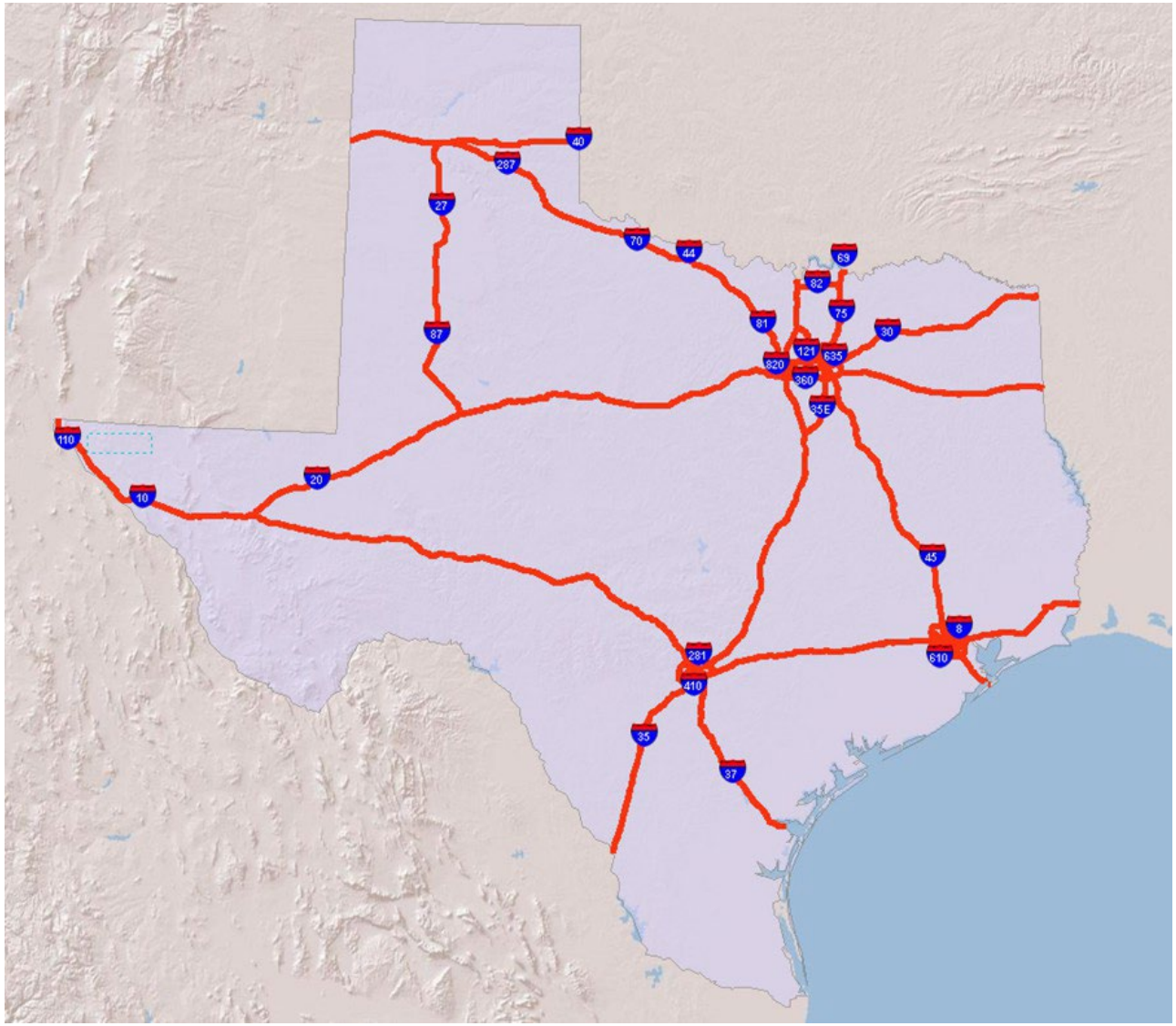
- **Coordinate pairs** form **points, lines, and polygons**



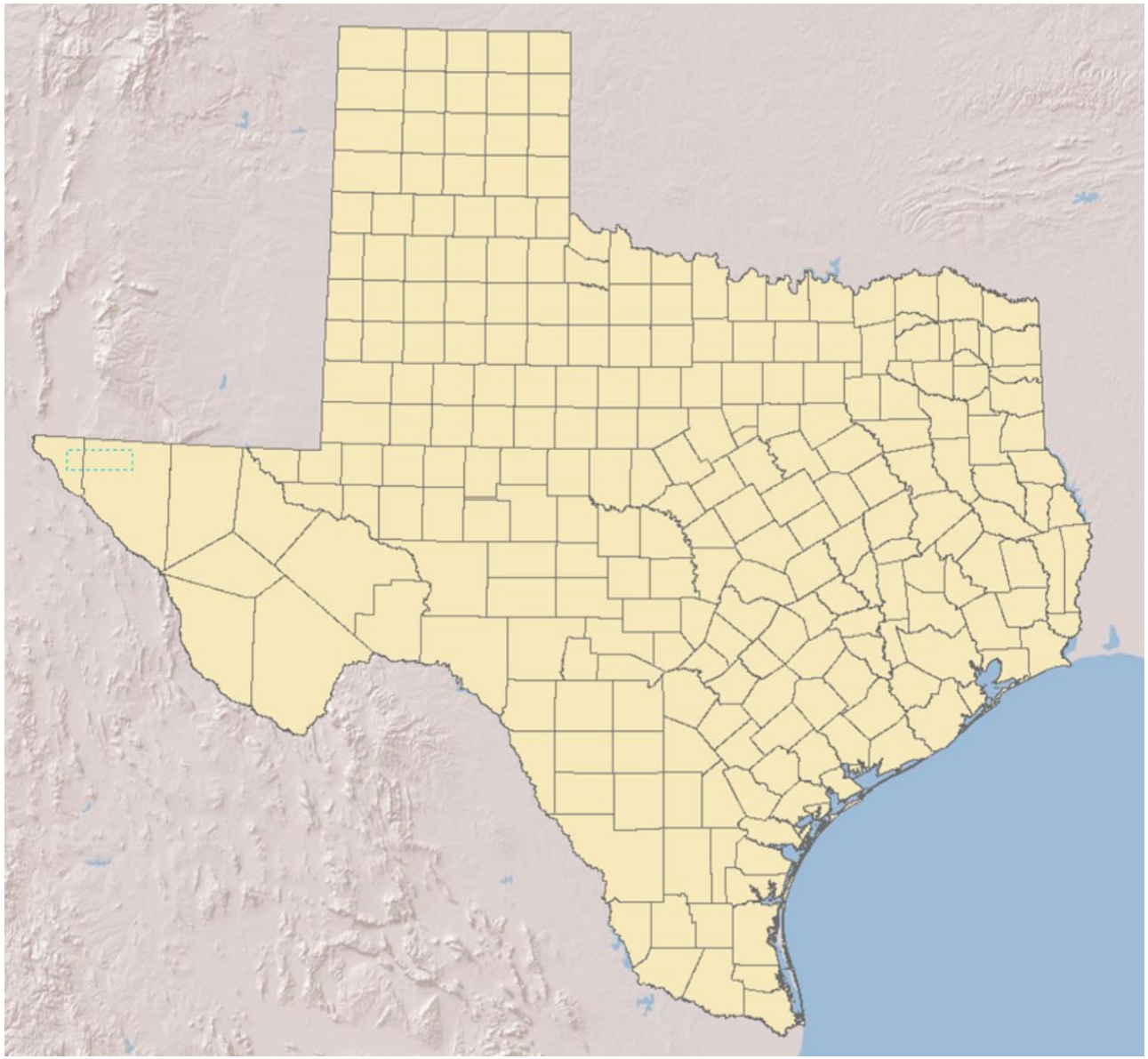
Points: Large cities in Texas



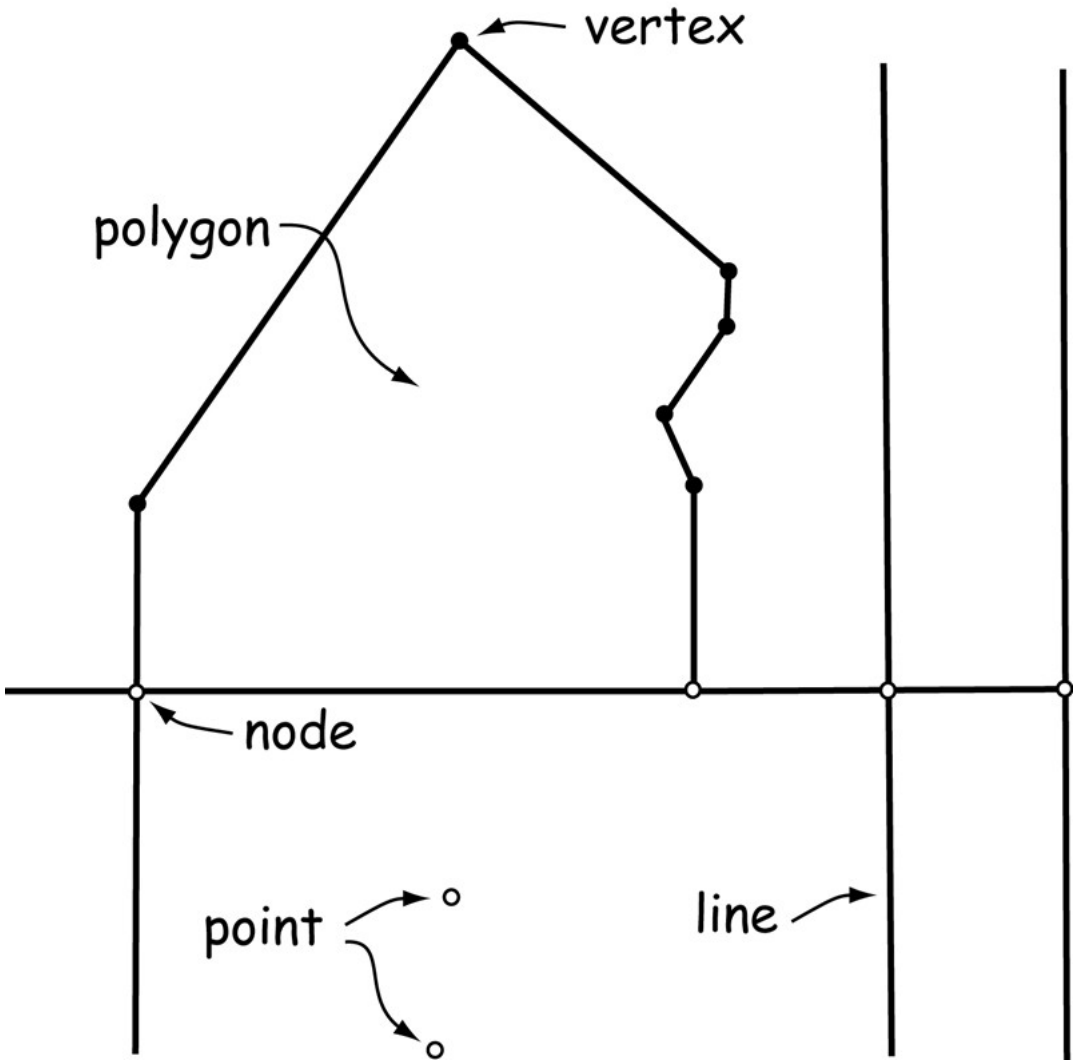
Lines: Interstate highways



Polygons: Counties



Vector data



Vector data: features and attributes

Geographic **feature** (point on the map) linked to corresponding **attributes** (in this case, demographic information)

NAME	POP2000	POP2007	WHITE	BLACK	AMERI_ES	ASIAN	HAWN_PI	OTHER	MULT_RACE	HISPANIC	MALES	FEMALES
El Paso	563662	608849	413061	17586	4601	6321	583	102320	19190	431875	267651	296011
Arlington	332969	373933	225379	45727	1817	20015	475	29763	9793	60817	166465	166504
Fort Worth	534694	609154	319159	108310	3144	14105	341	75100	14535	159368	263720	270974
Dallas	1188580	1259459	604209	307957	6472	32118	590	204883	32351	422587	598991	589589
San Antonio	1144646	1259735	774708	78120	9584	17934	1067	221362	41871	671394	553245	591401
Austin	656562	736677	429100	65956	3889	30960	469	106538	19650	200579	337569	318993
Houston	1953631	2109413	962610	494496	8568	103694	1182	321603	61478	730865	975551	978080

Vector data in ArcGIS: the Shapefile

- Simple vector data format (spaghetti)
- Fast to display, readable by most GIs packages
- Composed of several related files
    - Required: **.shp** (feature geometry), **.dbf** (attributes), **.shx** (index file)
    - Optional, but recommended: **.prj** (coordinate system information), **.sbn** and **.sbx** (spatial index files), **.shp.xml** (metadata)

Multipart features

- Linear and polygon features can sometimes have multiple parts (e.g. Florida)

The GIS understands the hundreds of islands in Florida as one “feature” in the feature class

Table

US_state_2010

FID	Shape	REGION10	DIVISION10	STATEFP10	STATENS10	GEOID10	STUSPS10	NAME10	LSA
31	Polygon	4	9	06	01779778	06	CA	California	00
32	Polygon	4	8	08	01779779	08	CO	Colorado	00
16	Polygon	1	1	09	01779780	09	CT	Connecticut	00
17	Polygon	3	5	10	01779781	10	DE	Delaware	00
18	Polygon	3	5	11	01702382	11	DC	District of Columbia	00
4	Polygon	3	5	12	00294478	12	FL	Florida	00
5	Polygon	3	5	13	01705317	13	GA	Georgia	00
11	Polygon	4	8	16	01779783	16	ID	Idaho	00
38	Polygon	2	3	17	01779784	17	IL	Illinois	00
39	Polygon	2	3	18	00448508	18	IN	Indiana	00
44	Polygon	2	4	19	01779785	19	IA	Iowa	00
33	Polygon	2	4	20	00481813	20	KS	Kansas	00

0 (1 out of 49 Selected)

US_state_2010

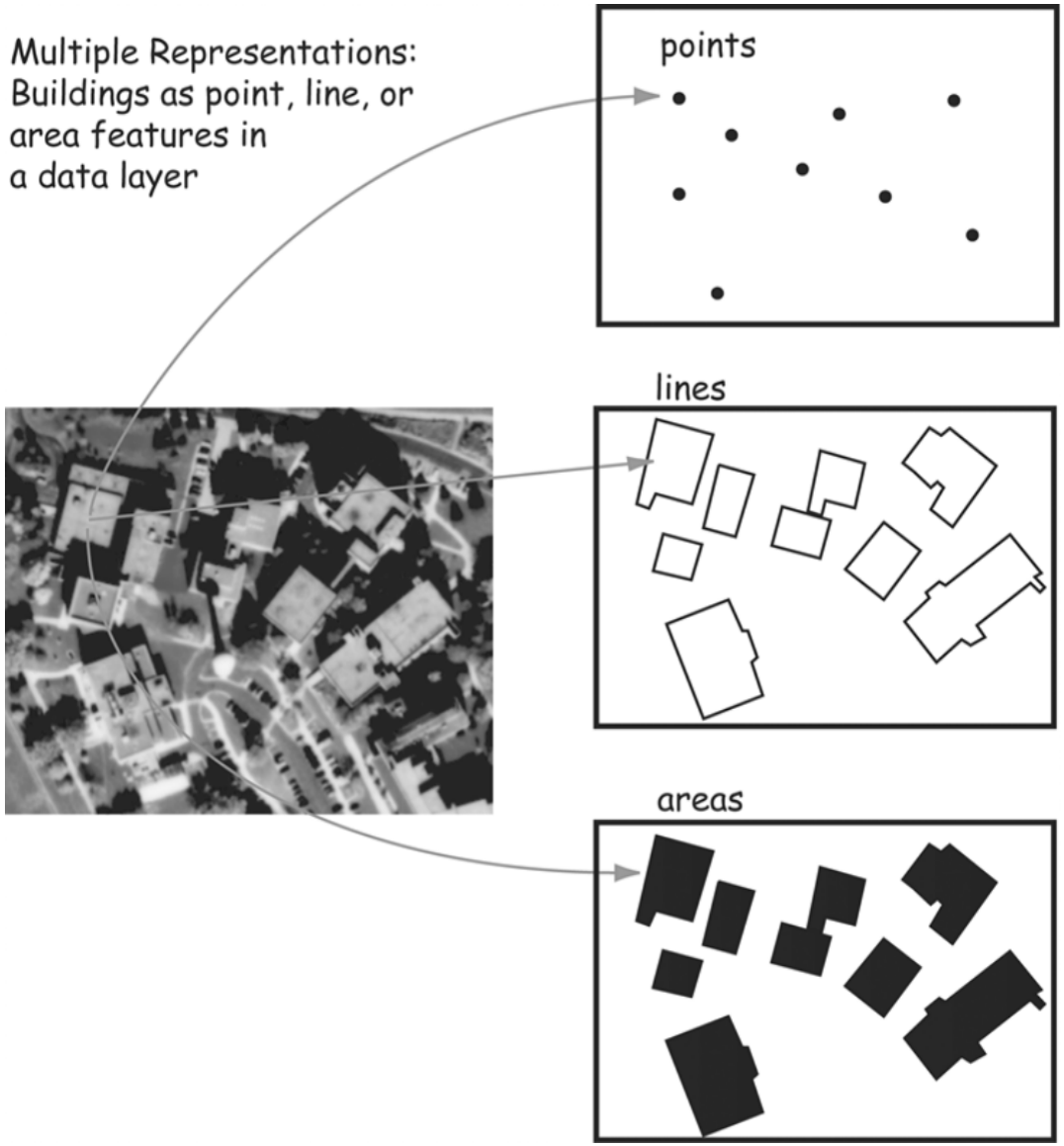
Polygon inclusion - what belongs?

Example: Key West, Florida

…whereas this area is not

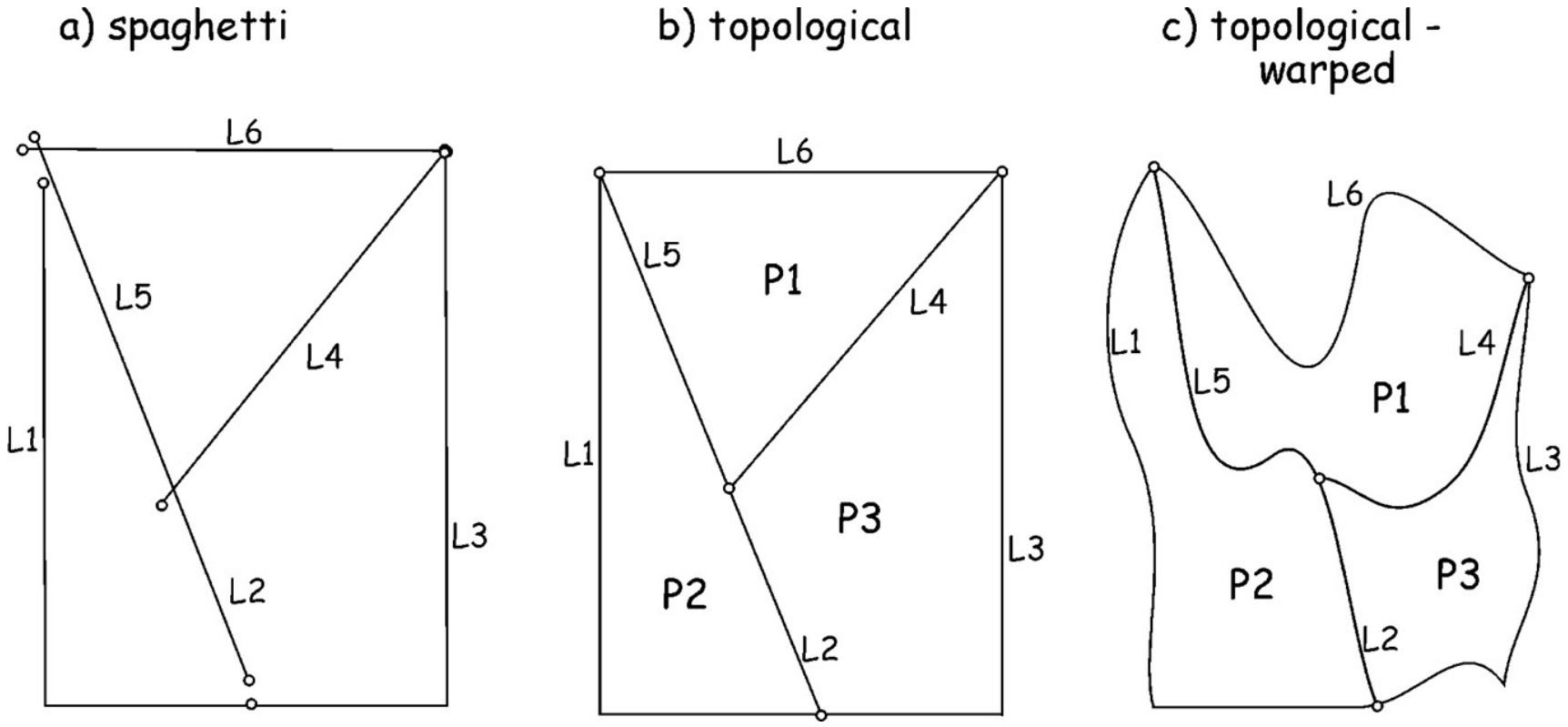
This water area is included as part of the polygon feature…

Point, line, polygon: which to choose?

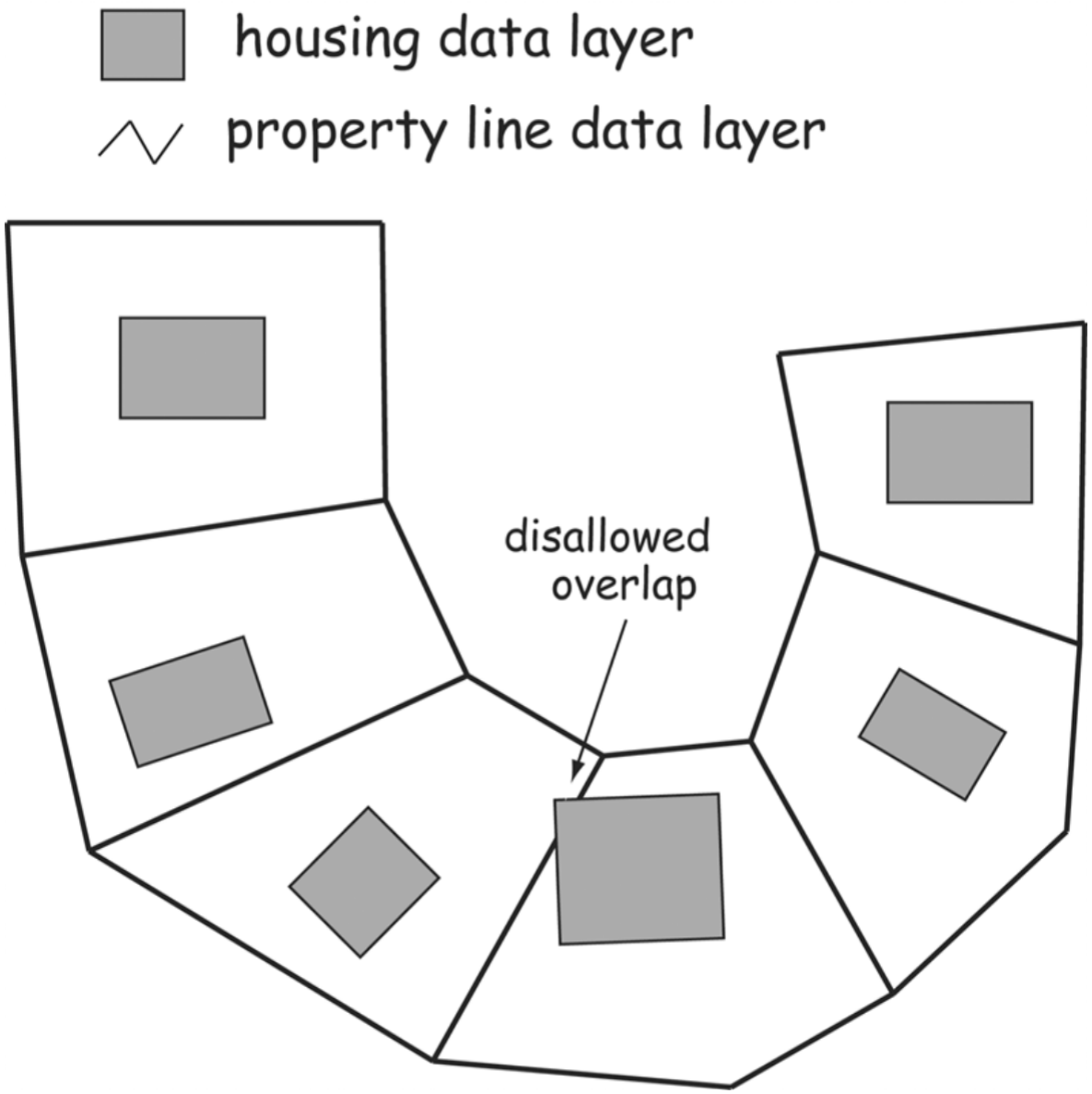


Topology and vector data

- Topology: rules governing **adjacency** and **connectivity** between geographic features



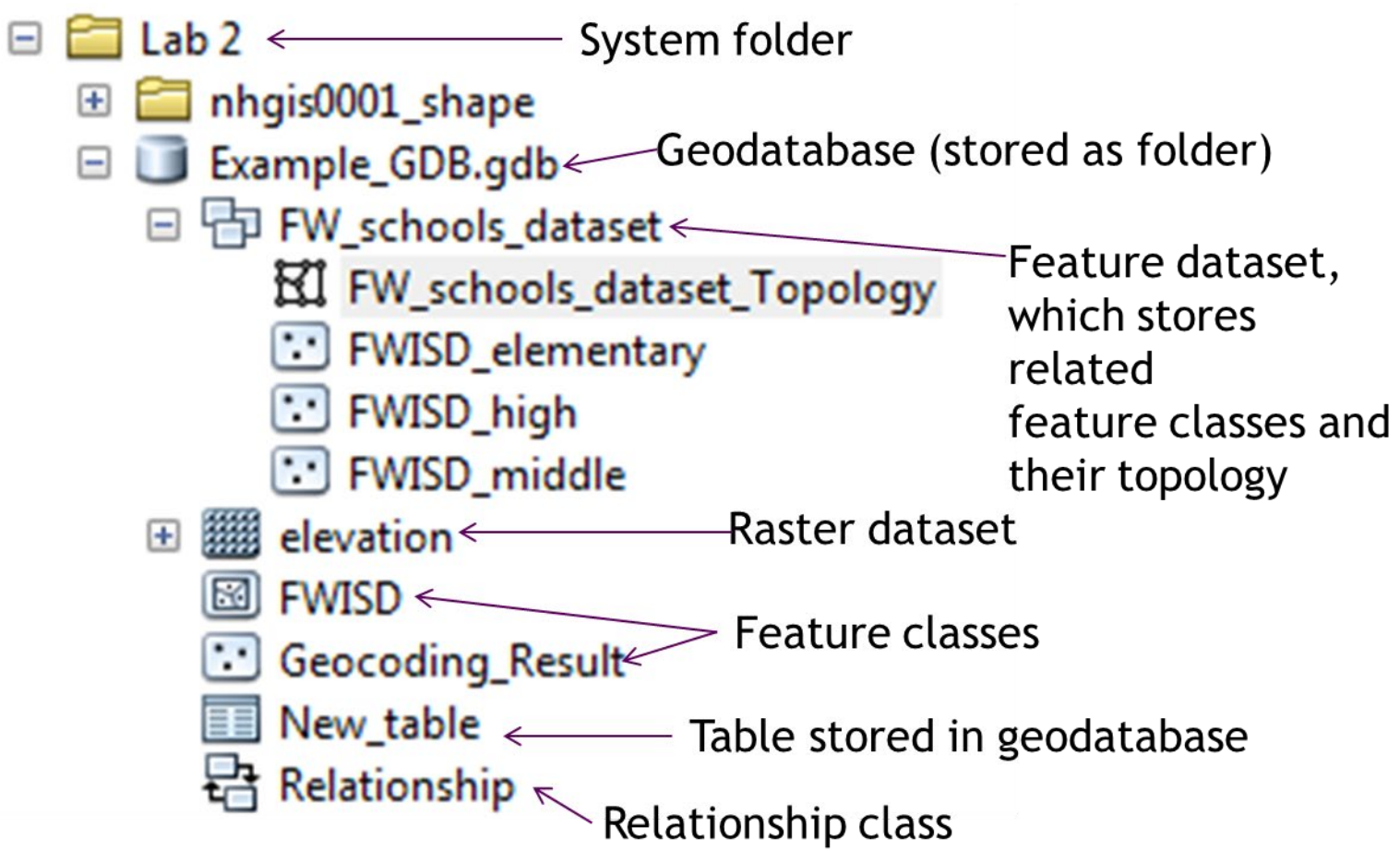
Topological rules



Vector data in ArcGIS: the geodatabase

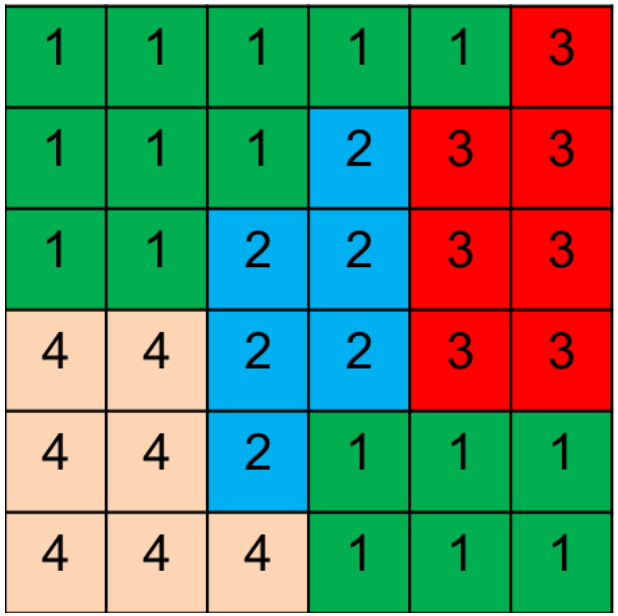
- Geodatabase: native data storage format for ArcGIS
- Stores vector data as tables; can also store relationships between tables, raster datasets, etc.
- Formats:
    - File geodatabase: collection of datasets held in a file system folder
    - Personal geodatabase: geographic data stored as a Microsoft Access database
    - ArcSDE (enterprise) geodatabase: collection of datasets stored in a relational database (for use on a network)

The Esri Geodatabase



Raster data

- Grid of cells that form a continuous surface
- Each grid cell contains a value that represents some real-world phenomena (e.g. elevation, land use)
- Cell sizes are referred to as the **resolution** of the data



1: Forest

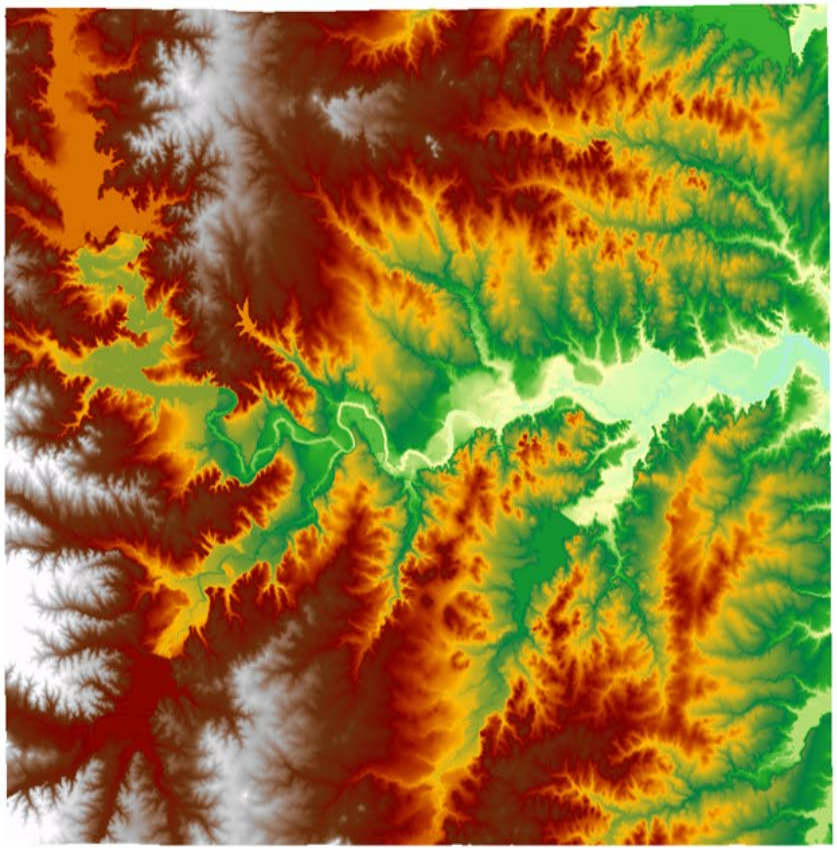
2: Water

3: Residential

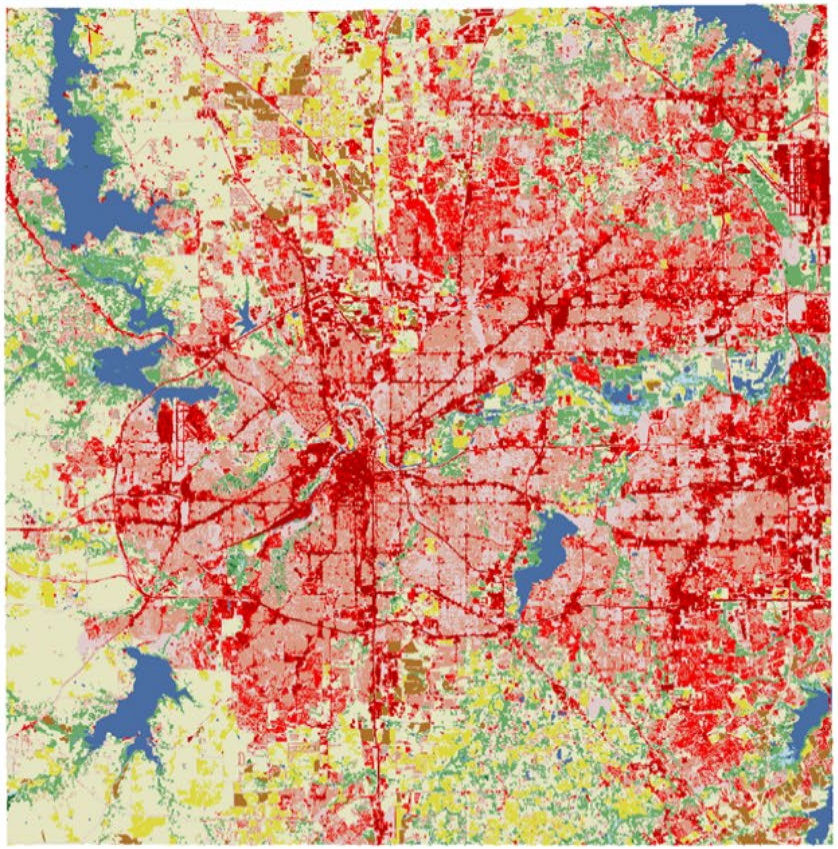
4: Agricultural

Quantitative and qualitative rasters

Elevation: quantitative



Land use: qualitative



Raster data resolution

- **Cell dimension:** the size of each grid cell (e.g. 30 meters by 30 meters)
- **Resolution:** the relative size of the grid cell

Low resolution

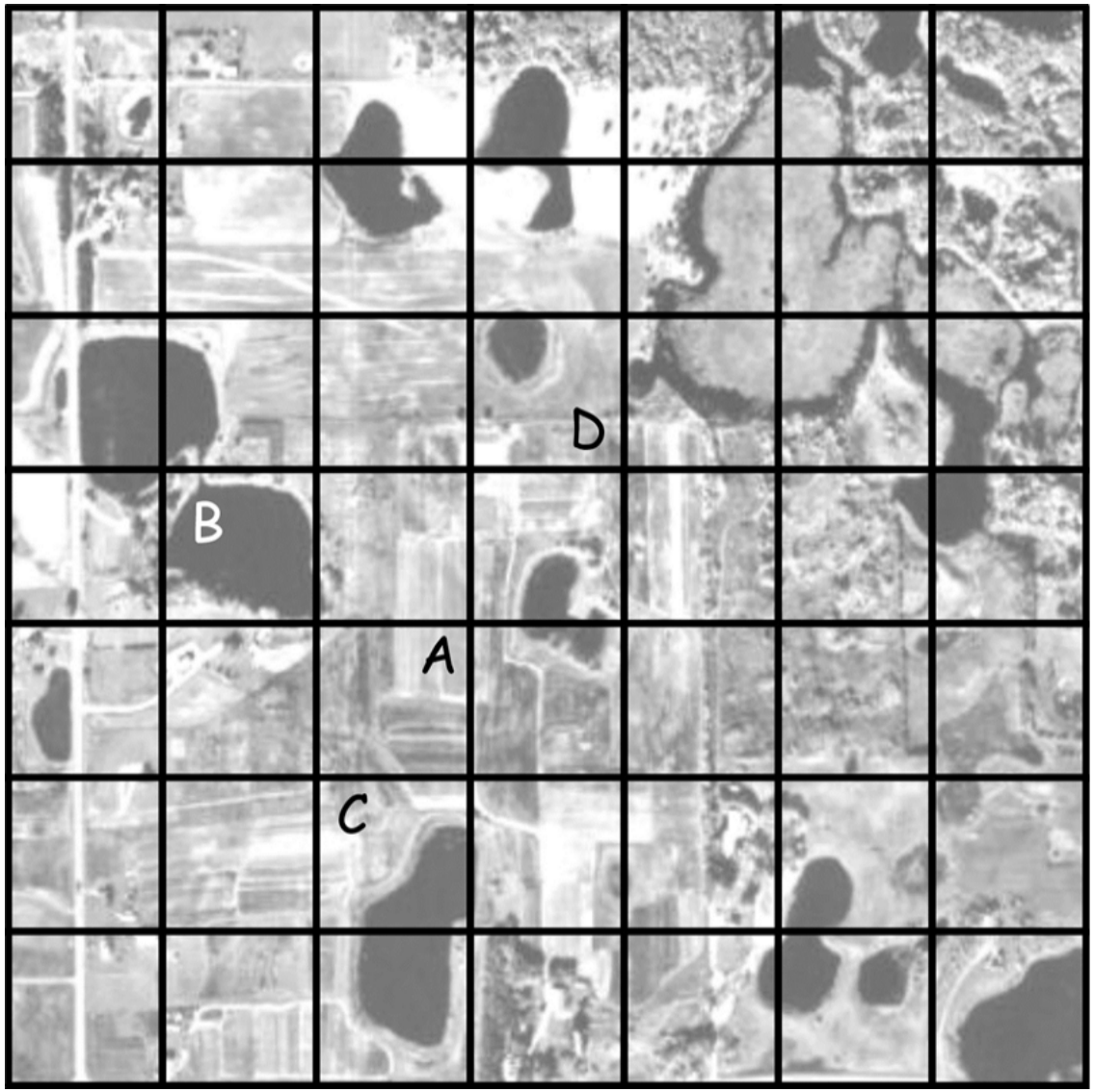
Scale 1:20,000
Cell size: 15 m

High resolution

Scale 1:20,000
Cell size: 15.24 cm

Image credit: ESRI

Raster data classification



How would you classify A, B, C, and D?

Raster data storage

- High-resolution raster datasets can contain millions of cells. and can get very large!
- Compression solutions: run-length encoding, quad tree representation

Run-length encoding

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

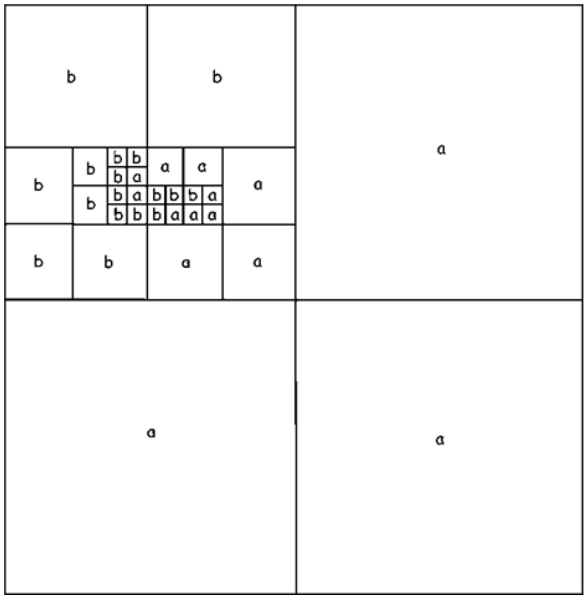
2:9, 5:6, 1:7

8:6

2:9, 4:6, 2:7

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

Quad-tree representation



Raster vs. vector

- What might be some advantages or disadvantages of each model?

Raster vs. vector

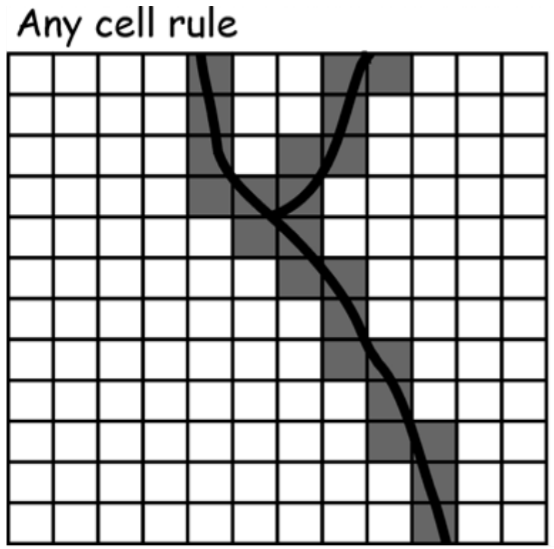
Advantages of raster data:

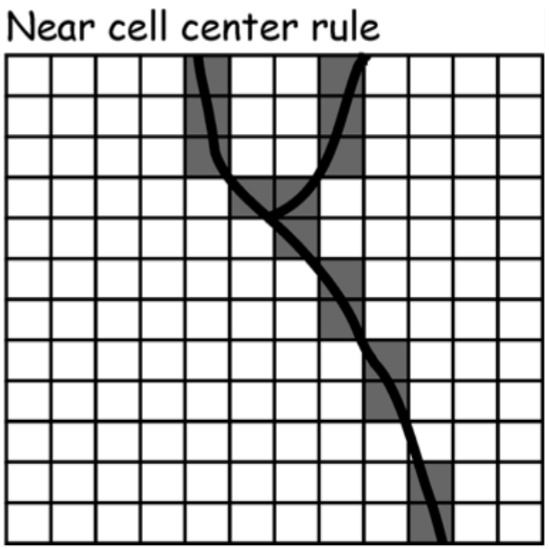
- Simpler data structure
- Straightforward data analysis (especially for continuous phenomena)
- Best for representing imagery

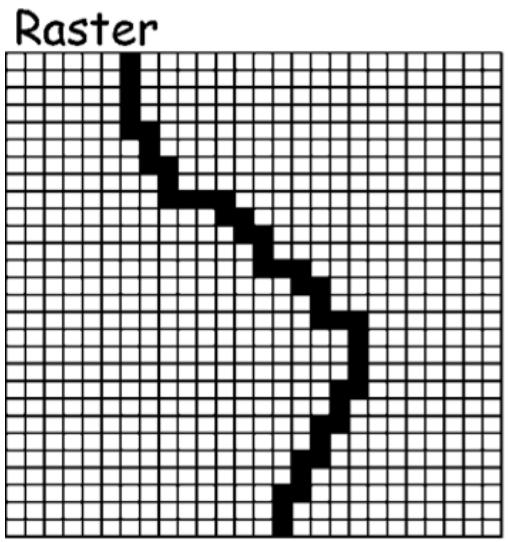
Advantages of vector data:

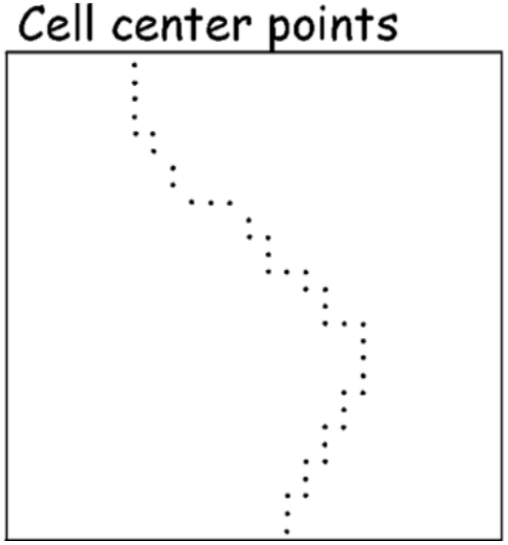
- Smaller file sizes
- Easy to re-project
- Good for some types of analysis (e.g. network analysis)
- Easy for map readers to understand

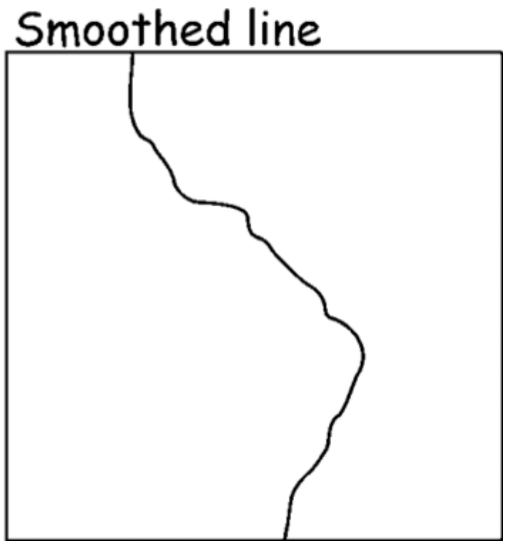
Vector/raster conversion





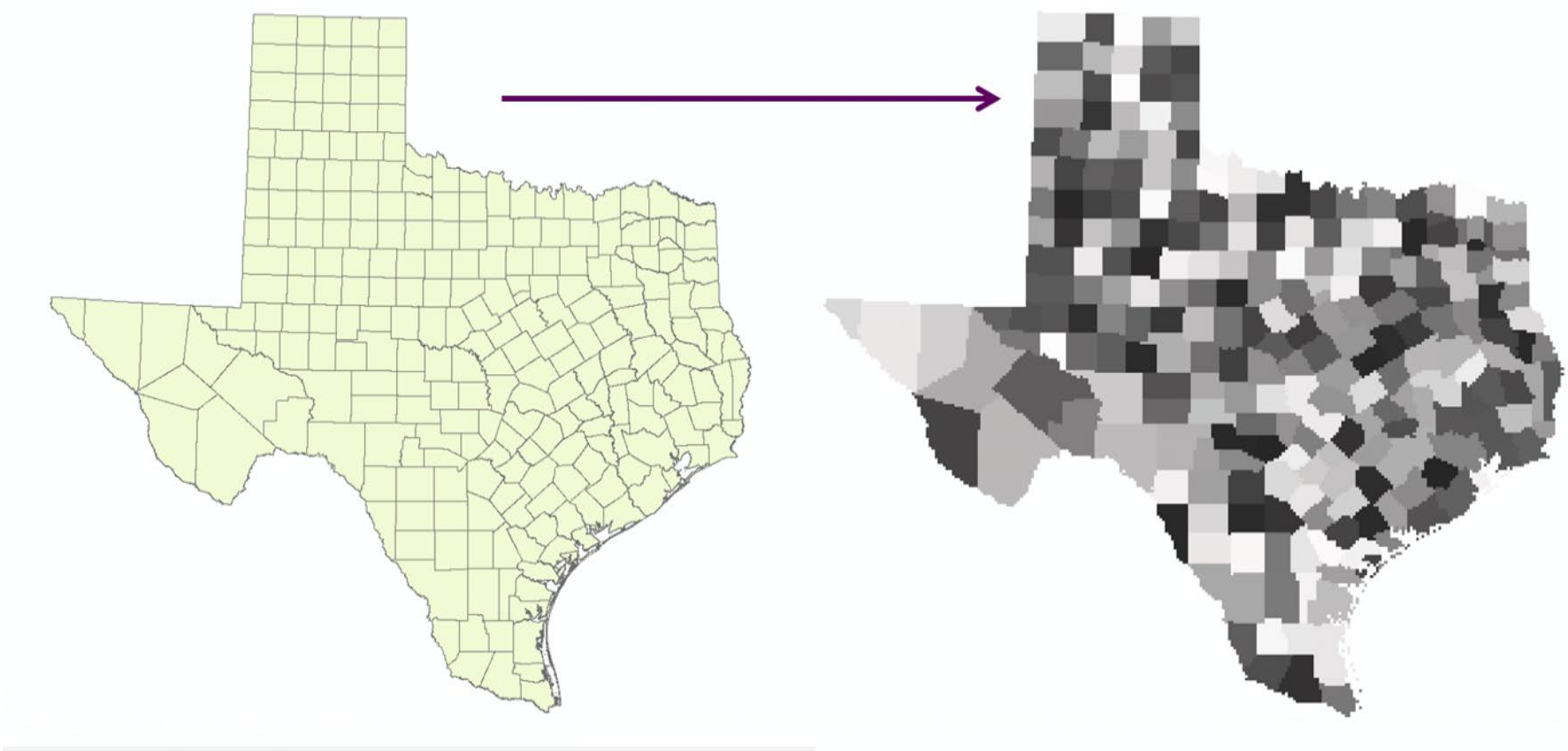




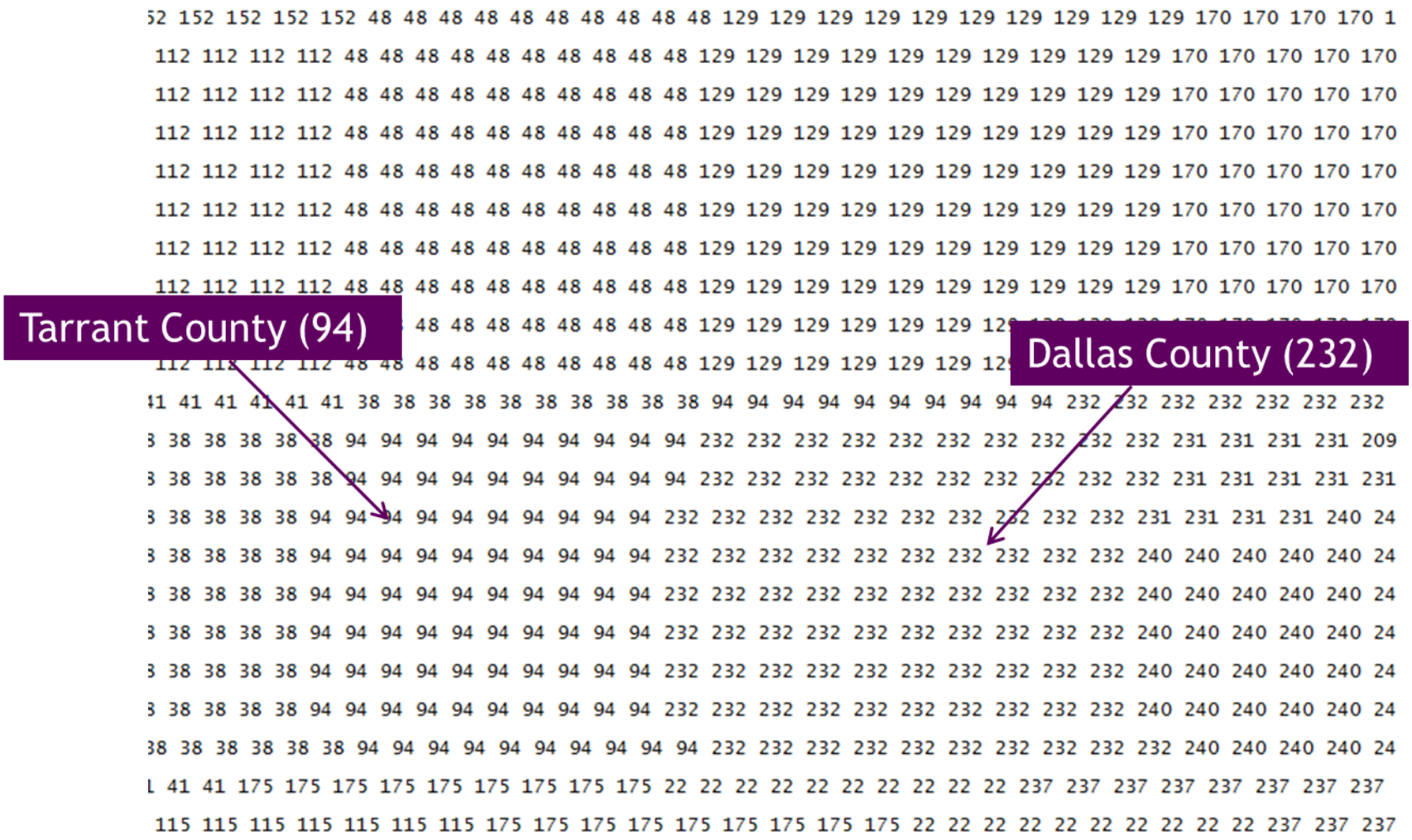


Does it make sense to convert?

- Example: counties in Texas

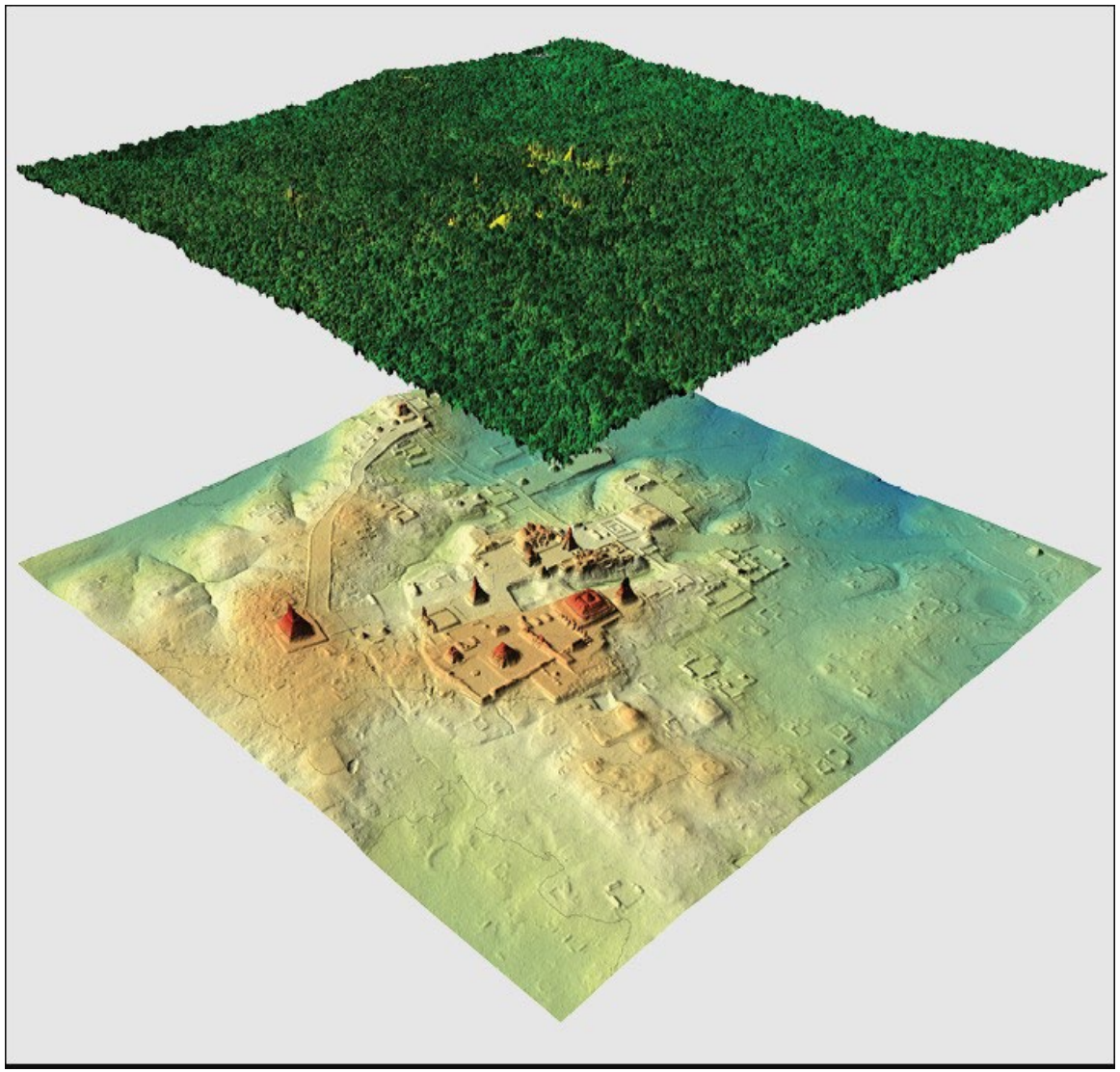


Texas counties as a raster



Other data types

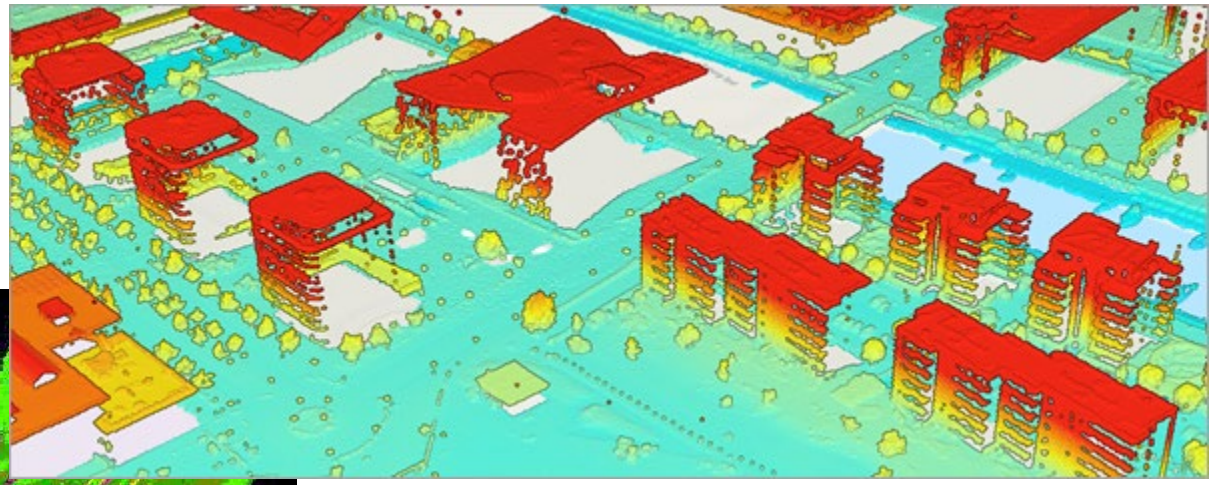
- LiDAR – Light detection and ranging technology creates “point cloud” data.



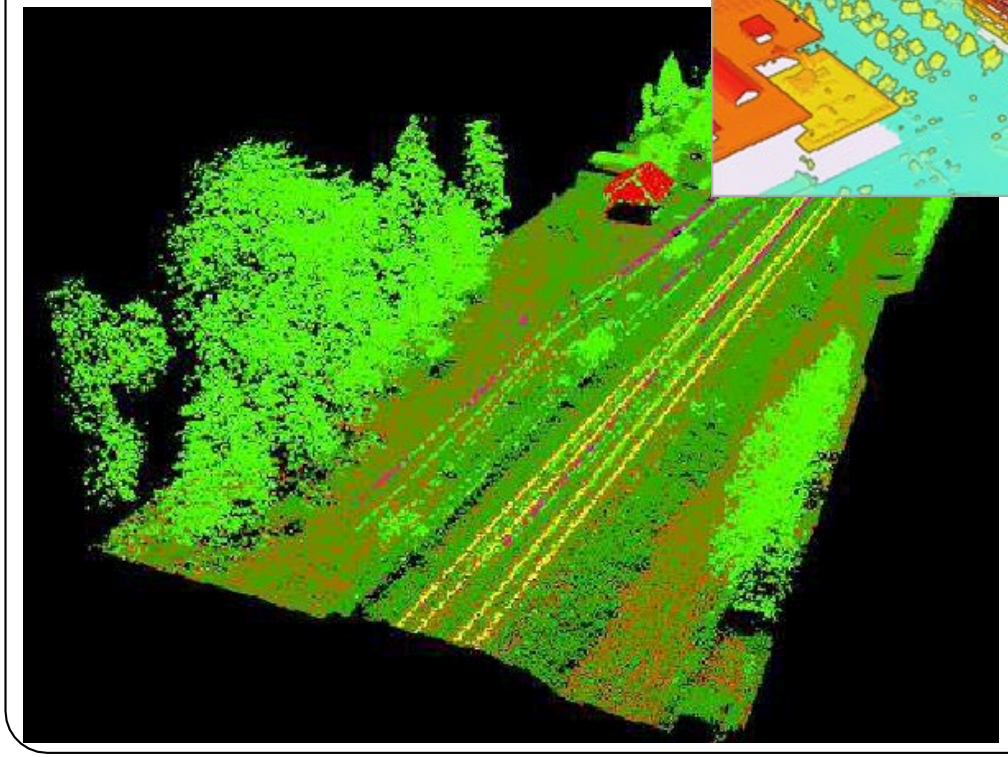
LiDAR Archeology

Other data types

- LiDAR – Light detection and ranging technology creates “point cloud” data.



Source



Source

Other data types

- LiDAR – Light detection and ranging technology creates “point cloud” data.
- CAD – Computer assisted drafting – generally used for indoor GIS. Engineering and architectural applications. Becoming more common in real-estate services firms.
- DEMs – Digital Elevation Models – file formats specifically designed for elevation modelling.

https://gisgeography.com/gis-formats/