Geospatial Data Analytics (GEOG 389)

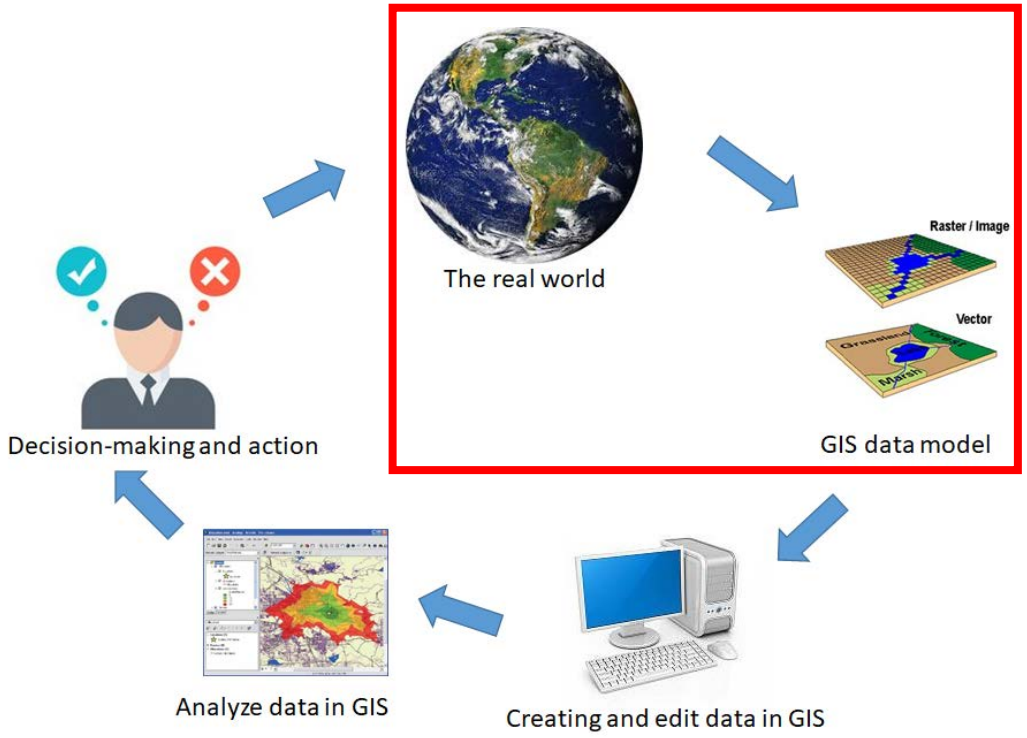
Spatial Data Models

Yi Qiang

February 24, 2019

Spatial Data Model

- A spatial data model is a simplified representation of ‘things’ on the earth surface.
- Spatial data models are designed for computer to process.
- Different spatial data models have different strengths in representing different types of ‘things’

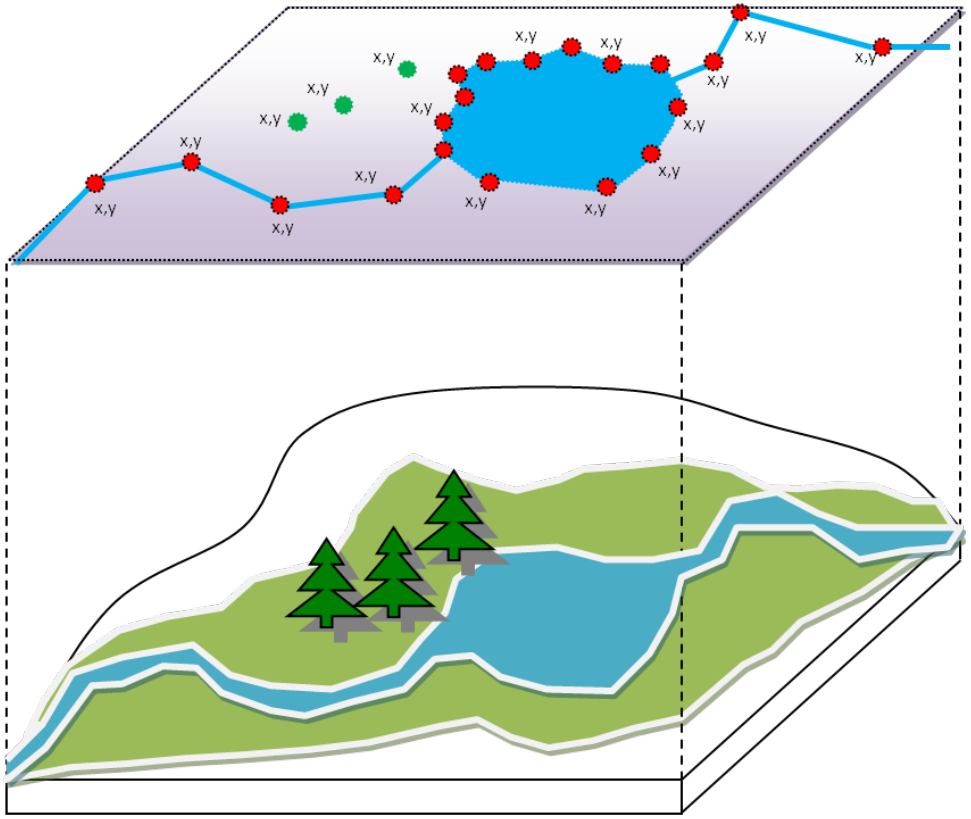


Representations of Space

- In the digital world, the two fundamental ways of representing space are **discrete object** and **continuous field**.
- **Discrete object**: the world is empty, except where it is occupied by objects with well-defined boundaries that are instances of generally recognized categories, for instance, a land parcel, a lake, a park or a building on the Earth surface.
- **Continuous field:** the world is filled values of one or multiple variables (e.g. elevation, temperature, population density). The variable(s) have values at every position.
- Discrete objects are represented in **vector model**, while continuous field is represented in **raster model**.

Vector Model

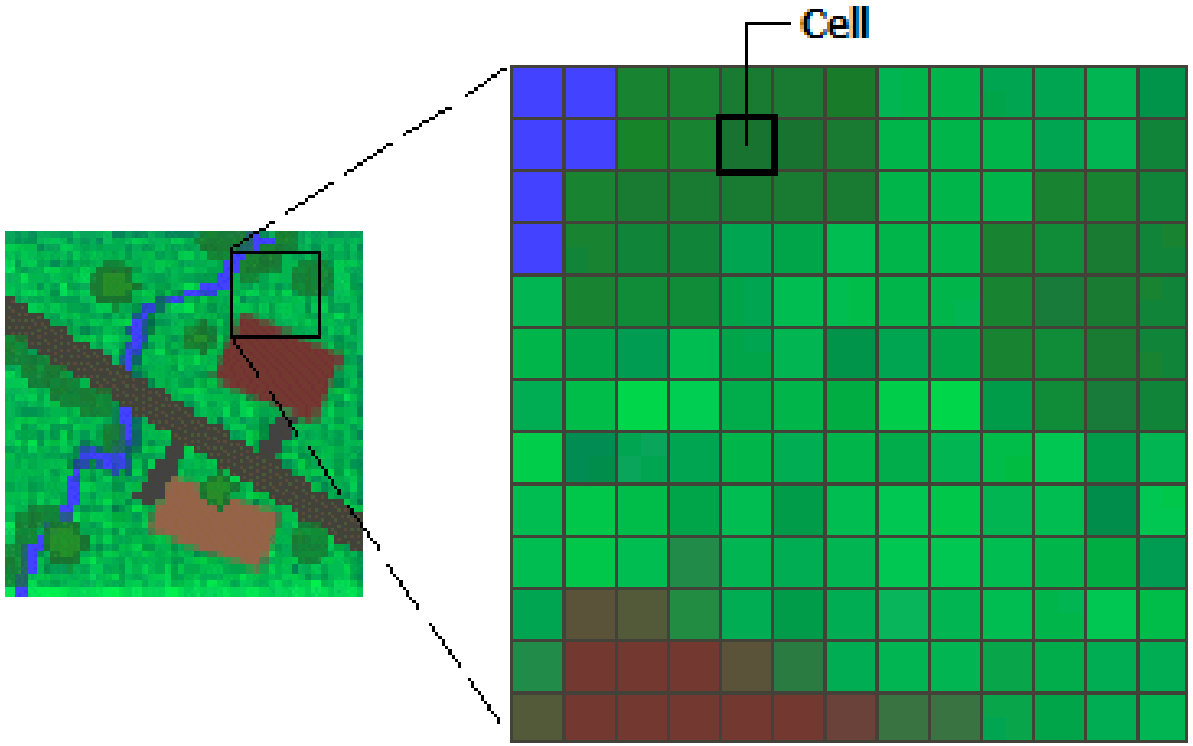
- Vector data use points, lines and polygons to represent spatial features as discrete objects.
- Good at representing spatial features with clear boundary and even interior.



Tree: point
Road: line
Lake: polygon

Raster data

- Raster data describe the world as a regular set of cells in a grid pattern.
- The cells are typically squares and arranged in X, Y directions.
- Each cell is associated with a value represent the attribute (e.g. elevation, land cover, housing density…)



Vector or Raster ?

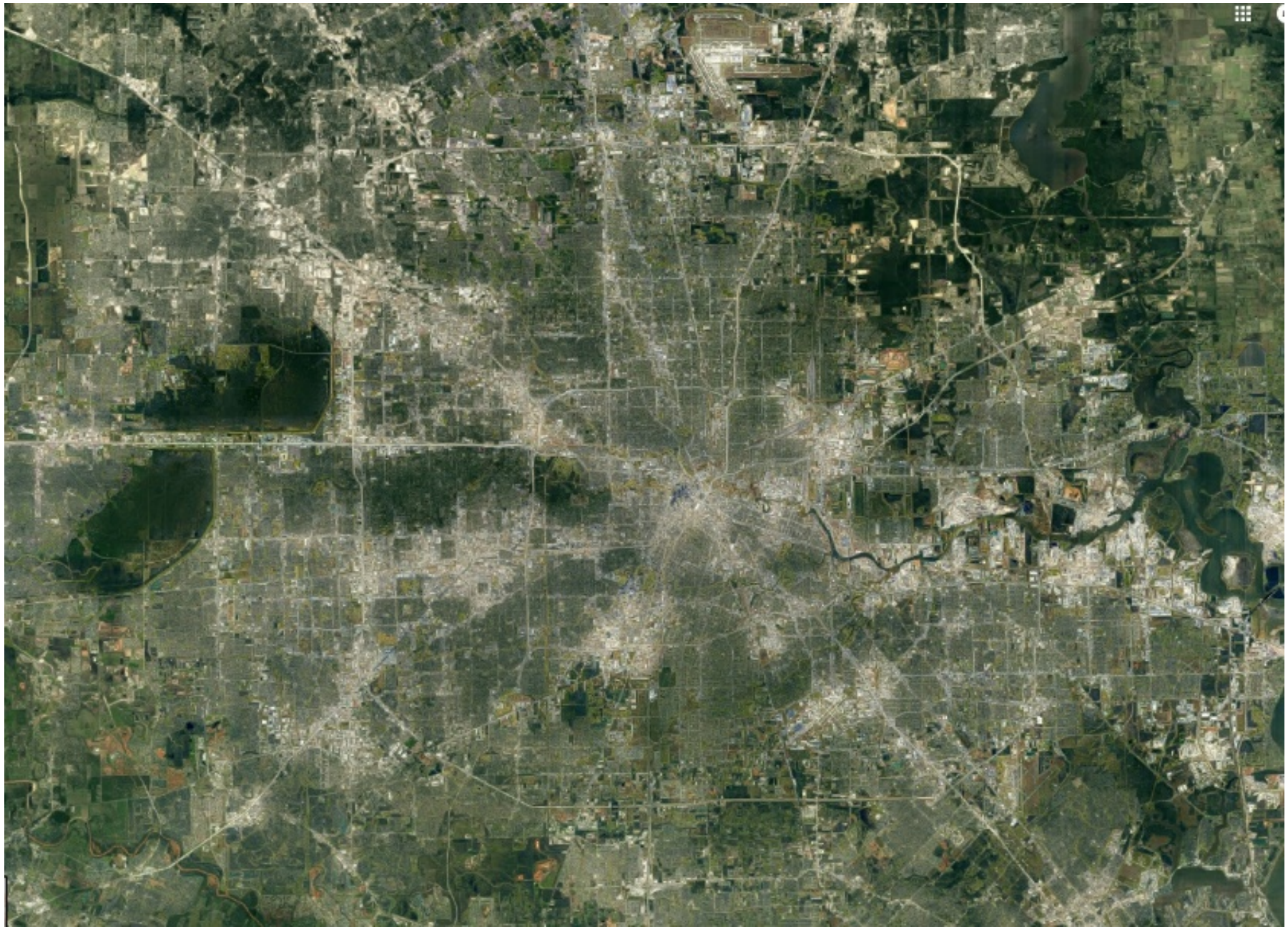
1. Lake
2. Road
3. Trees

Vector

Vector or Raster ?

1. Urban area
2. Green space

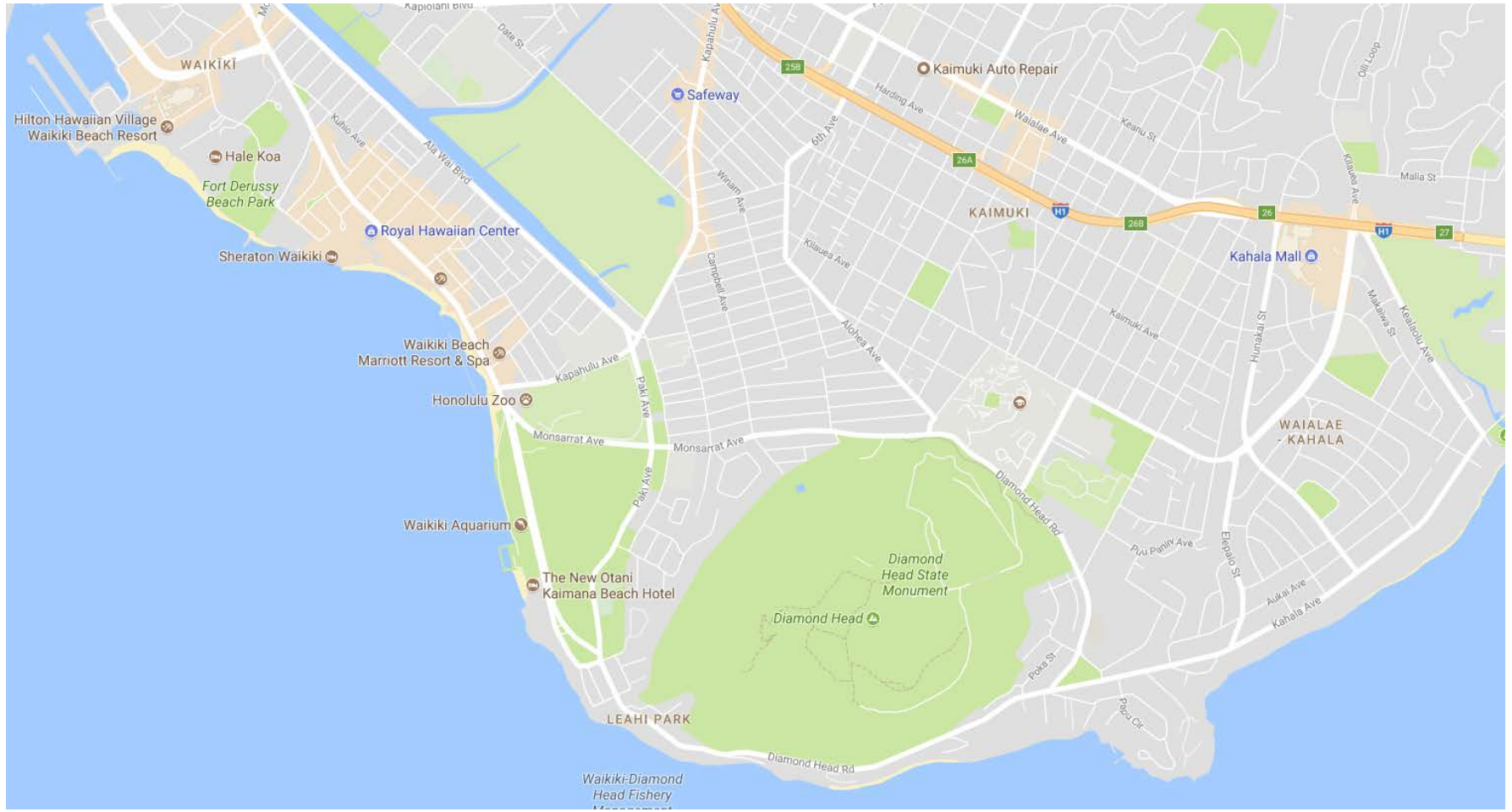
Raster



Vector or Raster ?

1. Vegetated land **Raster**
2. Road **Vector**

Raster (continuous fields)



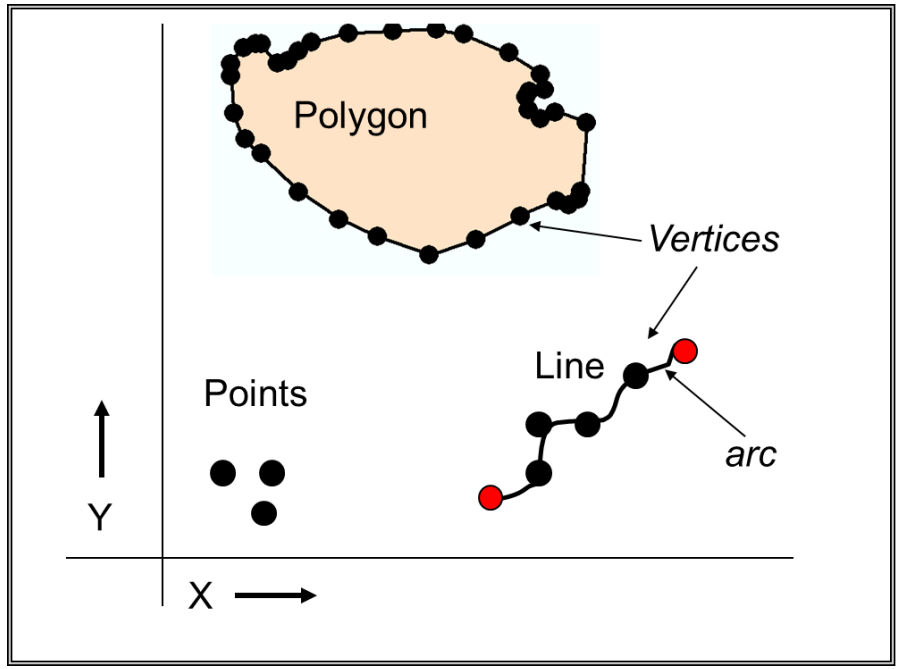
Vector (discrete objects)

Spatial representation is dependent on scales

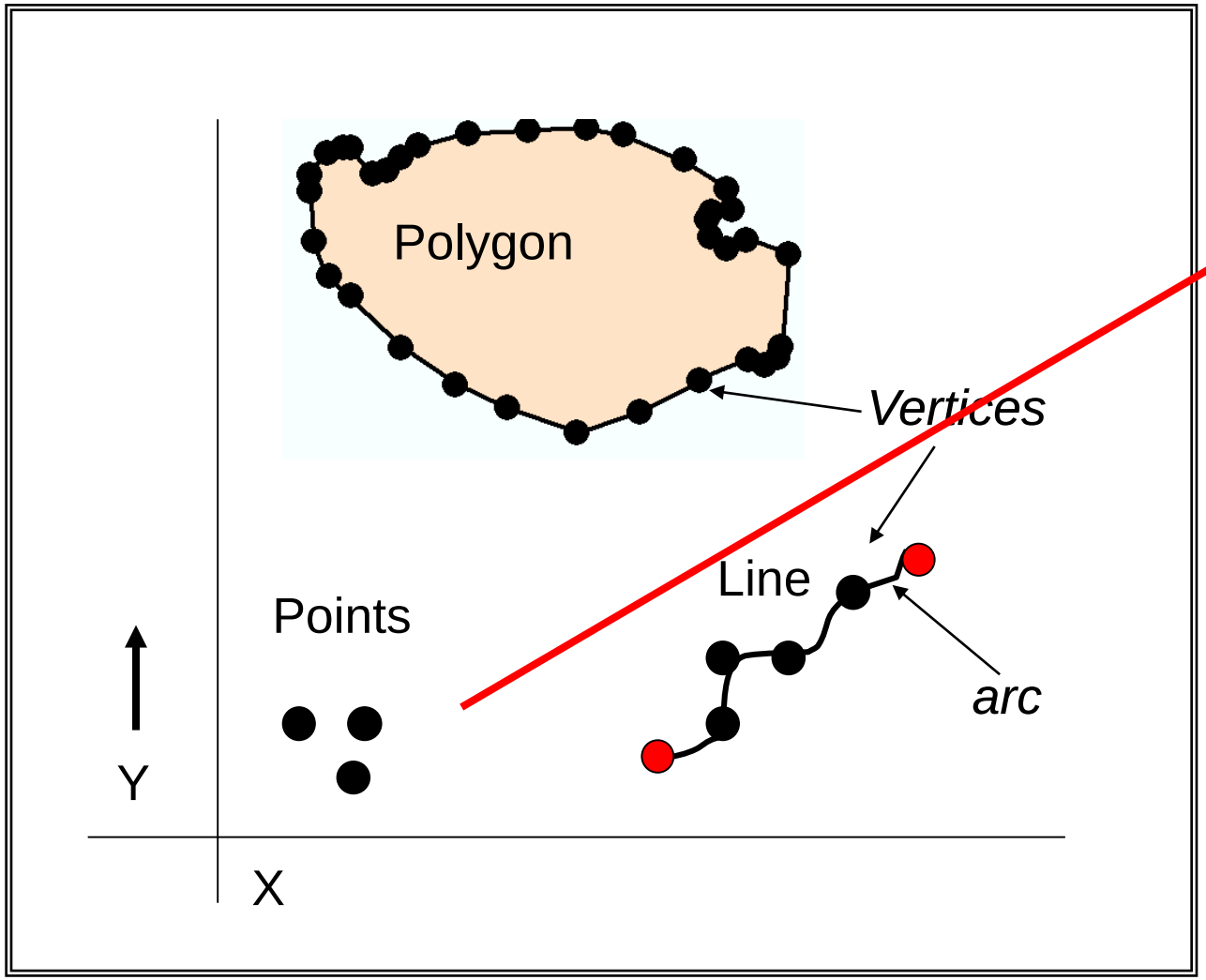
Where is the coast line?

Discrete Objects

- Spatial features are represented by three types of objects
  - **Points** as pairs of coordinates, in latitude/longitude, easting/northing or other coordinates
  - **Lines** as ordered sequence of points connected by straight lines
  - **Areas** as ordered rings of points, connected by straight lines or arcs to form **polygon**



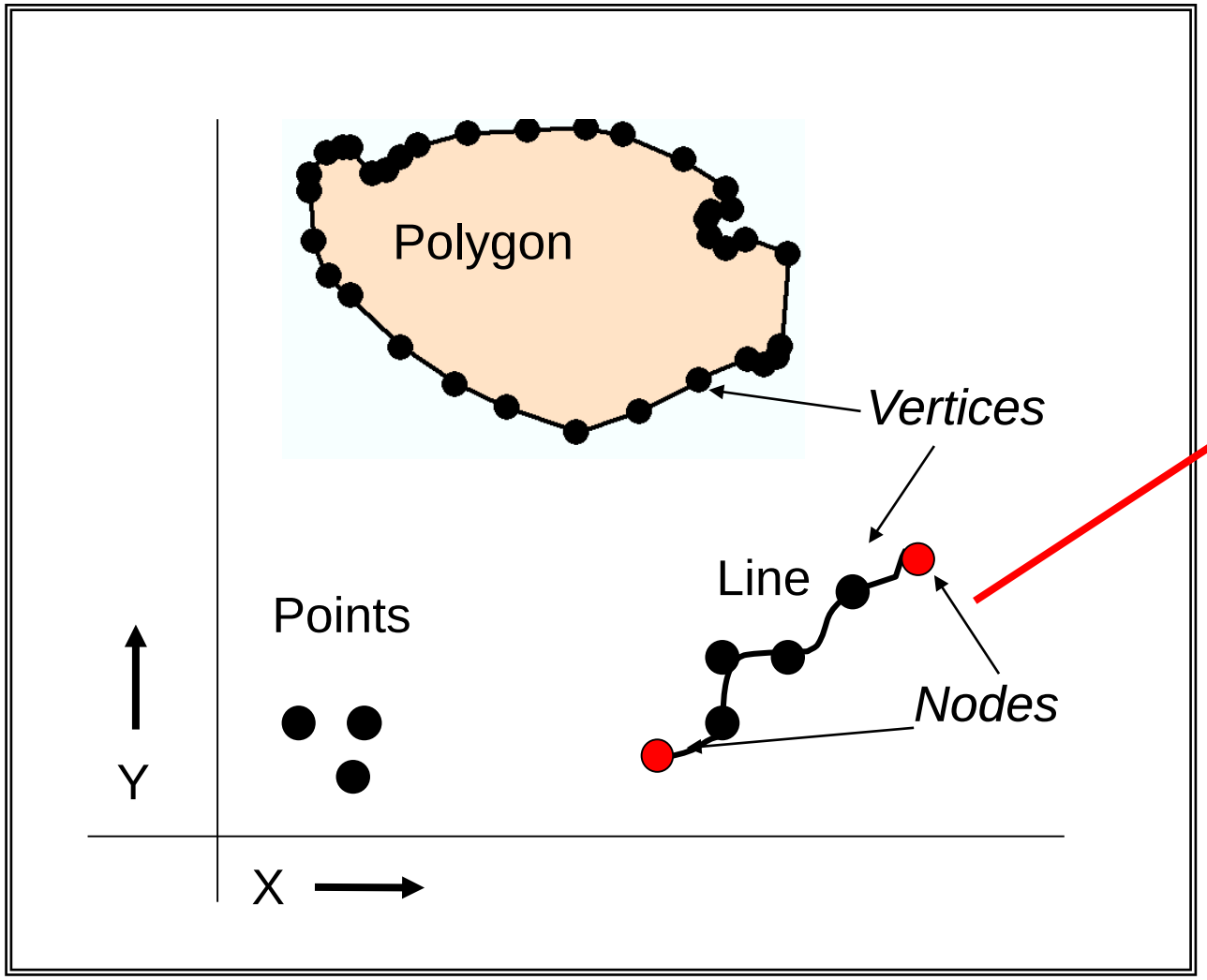
Encoding of discrete objects



Point 1: (X_1, Y_1)
Point 2: (X_2, Y_2)
Point 3: (X_3, Y_3)

X and Y are coordinates in a geographic coordinate system (degree/minute/second) or a projected coordinate system (meter, feet…)

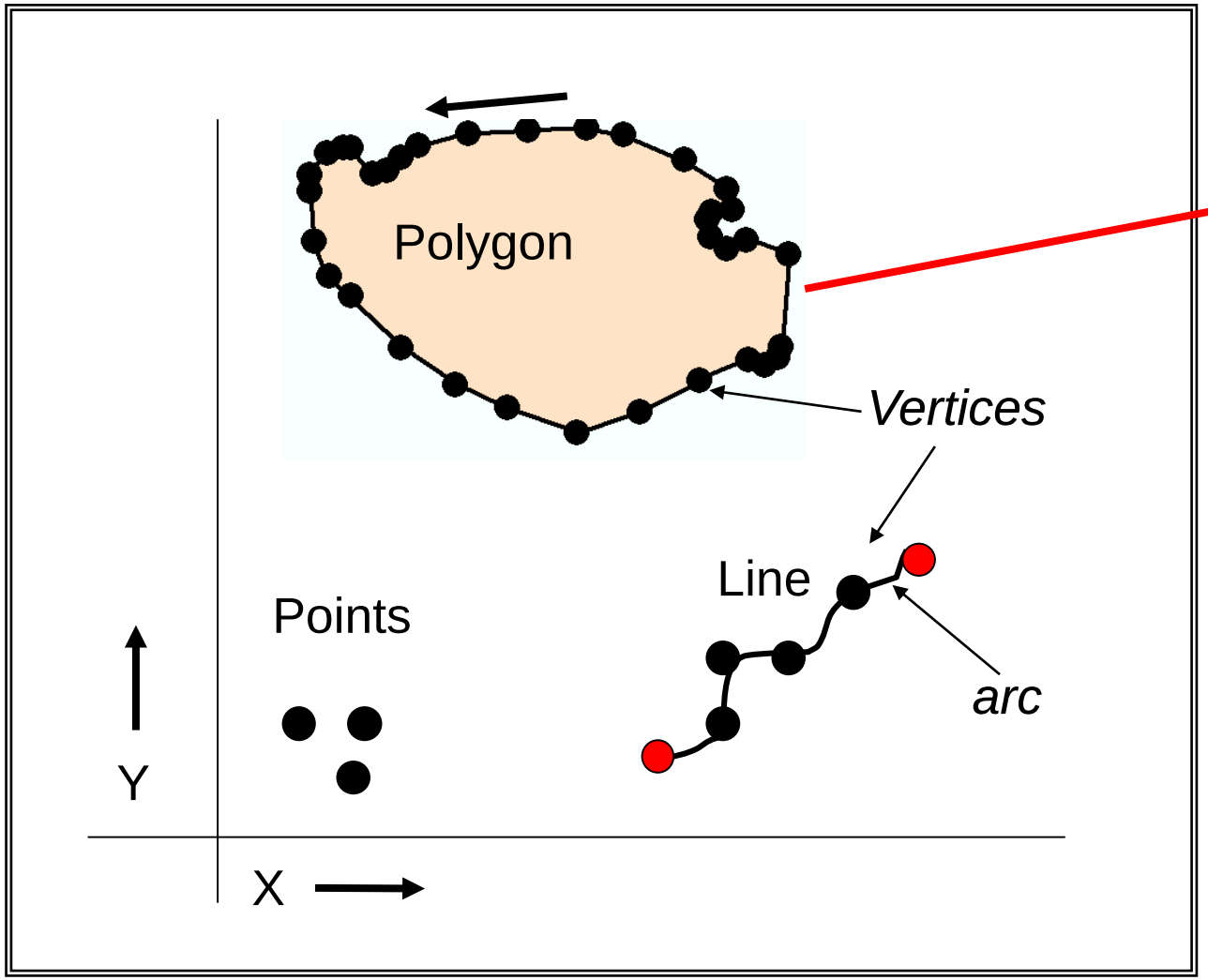
Encoding of objects



Line: $[(X_1, Y_1), (X_2, Y_2),\ldots, (X_n, Y_n)]$

- A sequence of points
- Adjacent points in the sequence are linked by straight-lines or arc.

Encoding of objects

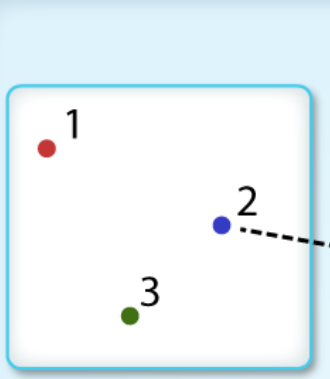


Polygon: $[(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)]$

- Defined by the boundary: sequence of points in a closed, non-self-intersecting loop
- Points are ordered counter-clockwise.

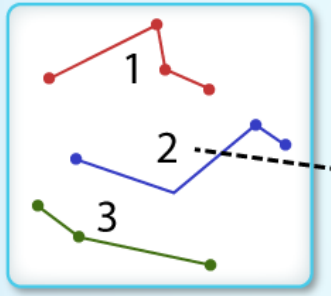
Attributes

- A place has characteristics or properties, which are called **attributes**. For example, the name of a place, the population of a city, the soil type of a farm.
- Each object is associated with one or a set of attributes by the unique ID.



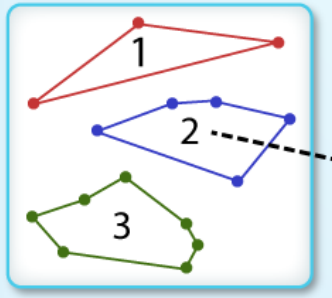
Example Attributes for Point Data

ID	Plot Size	Type	VegClass
1	40	Vegetation	Conifer
2	20	Vegetation	Deciduous
3	40	Vegetation	Conifer



Example Attributes for Line Data

ID	Type	Status	Maintenance
1	Road	Open	Year Round
2	Dirt Trail	Open	Summer
3	Road	Closed	Year Round

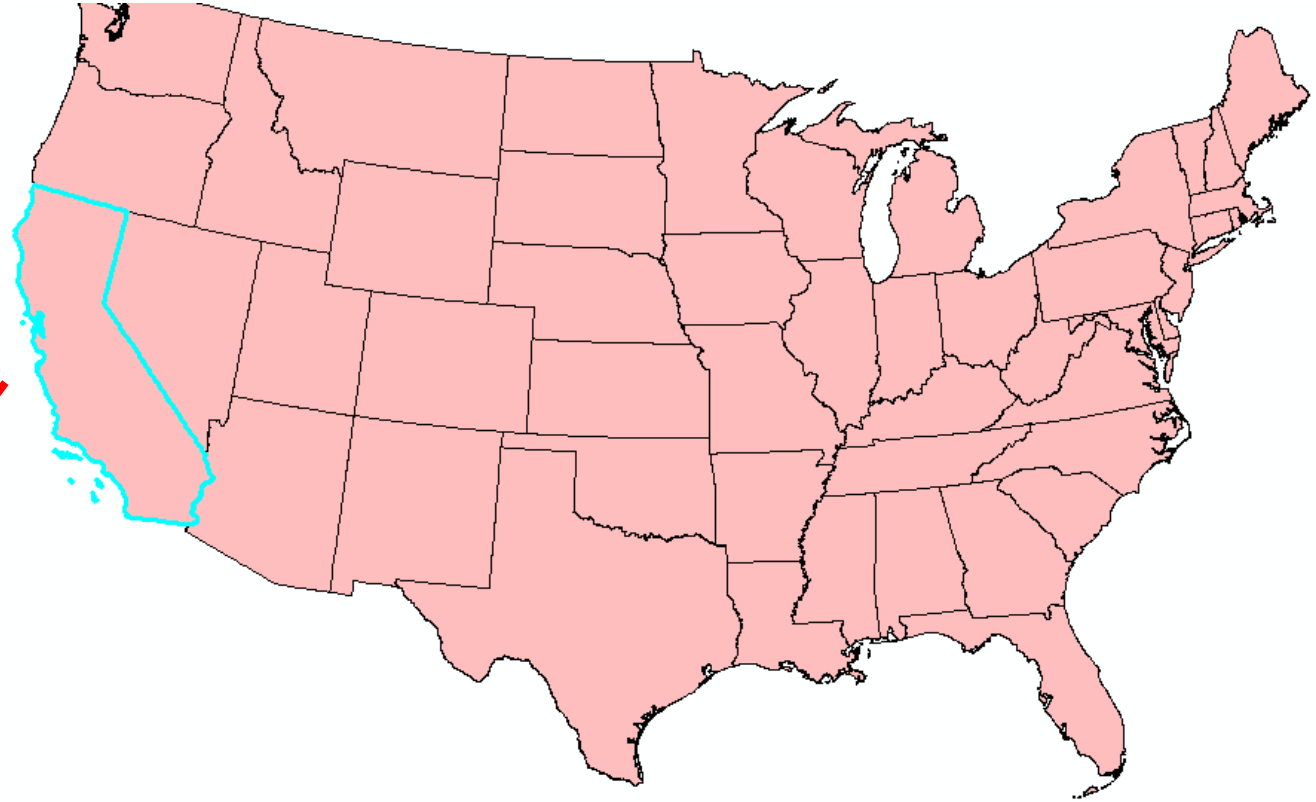


Example Attributes for Polygon Data

ID	Type	Class	Status
1	Herbaceous	Grassland	Protected
2	Herbaceous	Pasture	Open
3	Herbaceous / Woody	Grassland	Protected

Attributes

- Attributes are stored in a table, where each row represent the attributes of an object.
- The objects and their attributes are dynamically linked in a GIS.



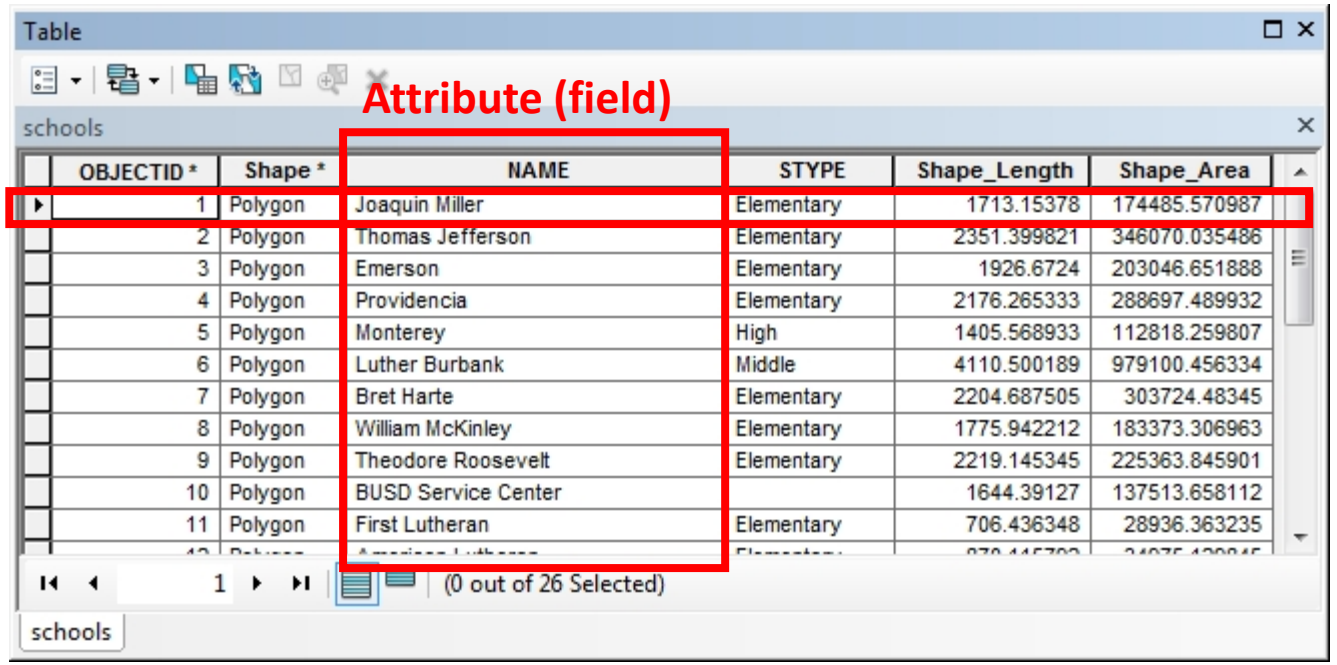
Table

cb_2016_us_state_500k

	FID	Shape	STATEFP	STATENS	AFFGEOID	GEOID	STUSPS	NAME
▸	0	Polygon	01	01779775	0400000US01	01	AL	Alabama
	1	Polygon	02	01785533	0400000US02	02	AK	Alaska
	2	Polygon	04	01779777	0400000US04	04	AZ	Arizona
	3	Polygon	05	00068085	0400000US05	05	AR	Arkansas
	4	Polygon	06	01779778	0400000US06	06	CA	California
	5	Polygon	08	01779779	0400000US08	08	CO	Colorado
	6	Polygon	09	01779780	0400000US09	09	CT	Connecticut
	7	Polygon	10	01779781	0400000US10	10	DE	Delaware
	8	Polygon	11	01702382	0400000US11	11	DC	District of Columbia
	9	Polygon	13	01705317	0400000US13	13	GA	Georgia
	10	Polygon	15	01779782	0400000US15	15	HI	Hawaii
	11	Polygon	16	01779783	0400000US16	16	ID	Idaho
	12	Polygon	17	01779784	0400000US17	17	IL	Illinois
	13	Polygon	18	00448508	0400000US18	18	IN	Indiana

Data Type in Attribute Table

- Each field (column) stores a particular data type.
- You **cannot** store different data types in the one field.
- Data type needs to be defined when you create the table/field – cannot change afterwards.



Data Type in Attribute Table

Common data types stored in attribute table:

- String (text): store characters, numbers and symbols
- Boolean: store binary values such as Yes/No, 1/0
- Numbers
  - Integer (long/short): store integer numbers without decimals
  - Float: store numbers with decimals (up to 7 digits)
  - Double: store long numbers with decimals (up to 16 digits)
- Time/Date

Data Formats of Vector Model

The Unified Structure

- Objects and attributes are stored in the same file, such as spreadsheets (csv, .xlsx), KML, GeoJSON
- Simple data structure, easy to edit and program
- More used in web-based applications

Object

Attributes

Feature ID	Geometry	Name	Country	Population
1	[21.3069, - 157.8583]	Honolulu	USA	351,792
2	[41.1579, 8.6291]	Porto	Portugal	214,349

[[21.3064, - 157.8582], [21.3059, -157.8673] , [21.3059, -157.8673] …]

Line/Polygon objects

EQ_MAG_	EQ_MAG_	EQ_MAG_	EQ_MAG_	EQ_MAG_	EQ_MAG_	INTENSITY	COUNTRY	STATE	LOCATION_NAME	LATITUDE	LONGITUDE	REGION_CODE	DEATHS	D
					7.3		JORDAN		JORDAN: BAB-A-DAR	31.1	35.5	140		
	7.1					10	TURKMENISTAN		TURKMENISTAN: W	38	58.2	40	1	
						10	SYRIA		SYRIA: UGARIT	35.683	35.8	130		
							GREECE		GREECE: THERA ISLAN	36.4	25.4	130		
						10	ISRAEL		ISRAEL: ARIHA (JERIC	31.5	35.3	140		
						10	ITALY		ITALY: LACUS CIMINI	35.5	25.5	130		
							SYRIA		SYRIAN COASTS	35.683	35.8	140		
					6.5		ISRAEL		ISRAEL: ARIHA (JERIC	32	35.5	140		
					6.2		JORDAN		JORDAN: SW: TIMNA	29.6	35	140		
						11	ISRAEL		ISRAEL: JERUSALEM	33	35.5	140		
							LEBANON		LEBANON: SUR (TYRE	33.27	35.22	140		
						10	GREECE		GREECE: MOUNT TAY	37	22.5	130		
							LEBANON		LEBANON: SUR (TYRE	33.56	35.37	140		
							GREECE		GREECE: SARONIC GU	37.9	23.5	130		
	7					9	GREECE		GREECE: MACEDONIA	39.7	23.3	130		
							GREECE		GREECE: ROMAN TER	37	22.5	130		
	7.1					10	GREECE		GREECE: EUBOEA	38.9	22.7	130	2500	
	7.6						IRAN		IRAN: REY,EIVAN-E-K	35.5	51.8	140		
	7.3					11	GREECE		GREECE	38.25	22.25	130		
							ITALY		ITALY: ROME	35	25	130		
	7					11	GREECE		GREECE: AEGEAN SEA	40	25	130		
							INDIA		INDIA: KUTCH	23	71	60		
						10	TURKEY		TURKEY: LYSIMACHIA	41	27	140		
							GREECE		GREECE: DELPHI	38.5	22.5	130		
	6.7					9	KYRGYZSTAN		KYRGYZSTAN: CHIGU	42.7	77.5	40		
	7.2					10	GREECE		GREECE: DODECANES	36.3	28.3	130		
							GREECE		GREECE			130		
						7	EGYPT		EGYPT: SIWA OASIS; I	29.2	25.5	15		
							SPAIN		SPAIN: CADIZ	36.2	-7.17	130		
							ITALY		ITALY: LIGURIA (LIGUI	44.3	8.5	130		
							PORTUGAL		PORTUGAL: CABO SA	36	-10.5	130		
							GREECE		EAST MEDITERRANEA	36.404	25.396	130		
	6.5						CHINA		CHINA: GANSU PROV	35.4	103.9	30		
	7						CHINA		CHINA: GANSU PROV	33.8	105.6	30	760	

Data Structure of Vector Model

The Unified Structure: GeoJSON

- JavaScript Object Notation (JSON)
- GeoJSON: a JSON format for geospatial data
- Based on a multi-level dictionary structure (key: value), e.g.

  In: data[‘geometry’][‘coordinates’]

  Out: [125.6, 10.1]

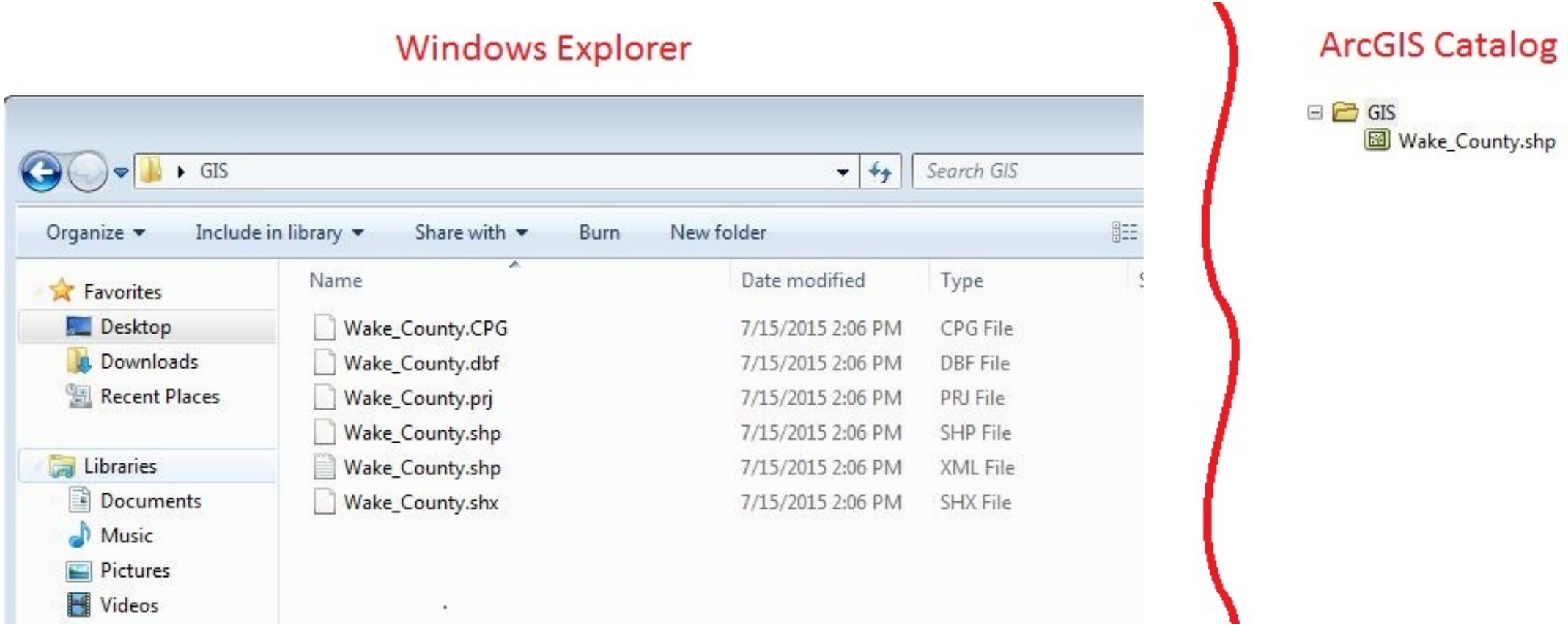
```
{
  "type": "Feature",
  "geometry": { "type": "Point", "coordinates": [125.6, 10.1]},
  "properties": {"name": "Dinagat Islands“, “pop": 125912
  }
}
```

A point feature in GeoJSON format

Data Structure of Vector Model

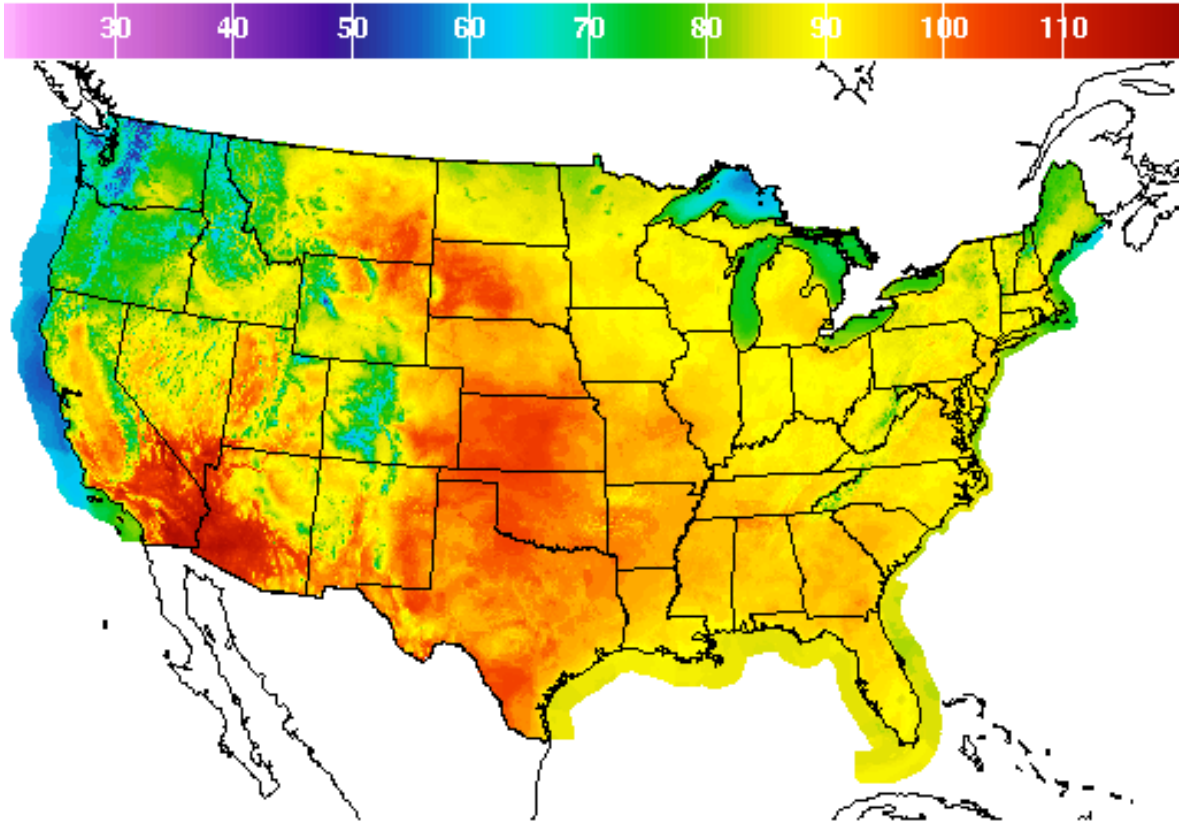
The Distributed Structure

- Objects and attributes are stored in separate files which linked by unique IDs, such as shapefile and geodatabase
- Fast processing and query in a desktop GIS



Continuous fields

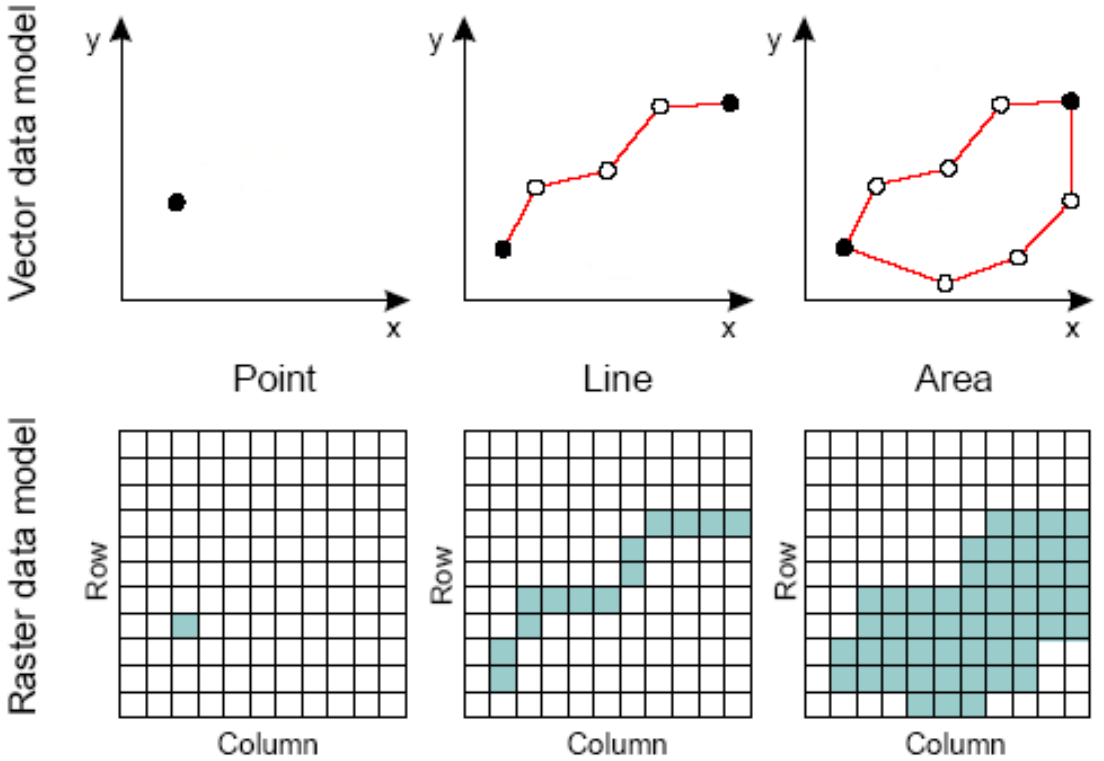
Discrete objects can represent spatial features with distinct shapes and boundaries. However, many geographic phenomena do not have clear shapes and boundaries, for instance, surface temperature, population density, elevation…



Continuous fields

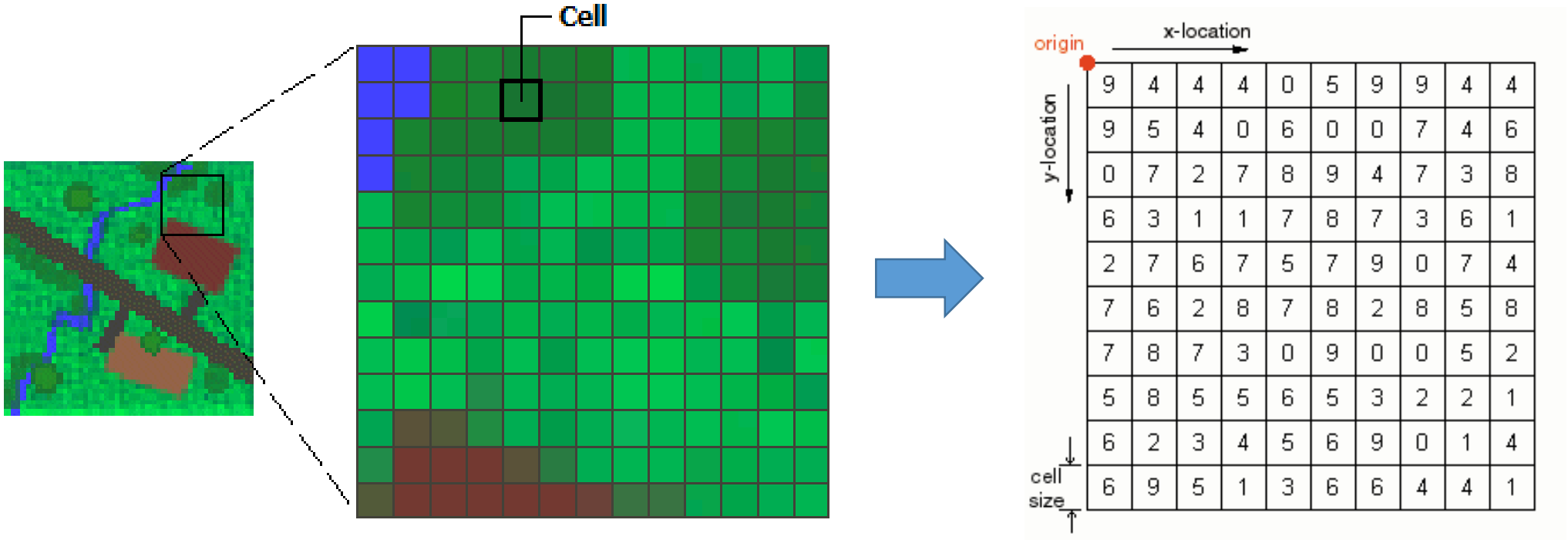
- **Continuous field:** the world is filled values of one or multiple variables (e.g. elevation, temperature, population density). The variable(s) have values at every position.
- Continuous field is represented in **raster model**, which is essentially a geolocated 2D matrix.

N	N	N	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N
N	N	1	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N



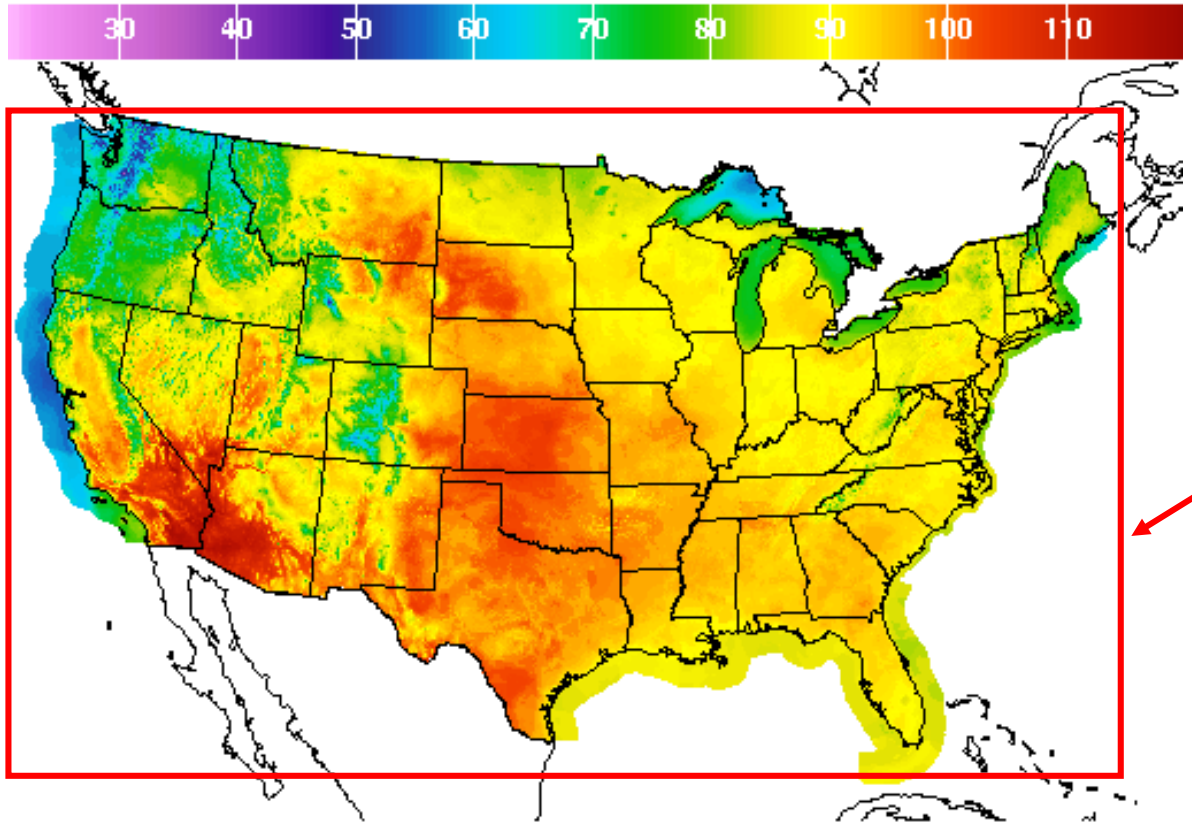
Raster Model

- Cells (also called pixels) in a raster are typically square and arranged in X (east), Y (north) directions.
- Each cell represent a specific area in the ground, and associated with values of the area (e.g. elevation, land cover, housing density…)
- Each raster must cover a ‘rectangular’ area.



Continuous fields

Discrete objects can represent spatial features with distinct shapes and boundaries. However, many geographic phenomena do not have clear shapes and boundaries, for instance, surface temperature, population density, elevation…



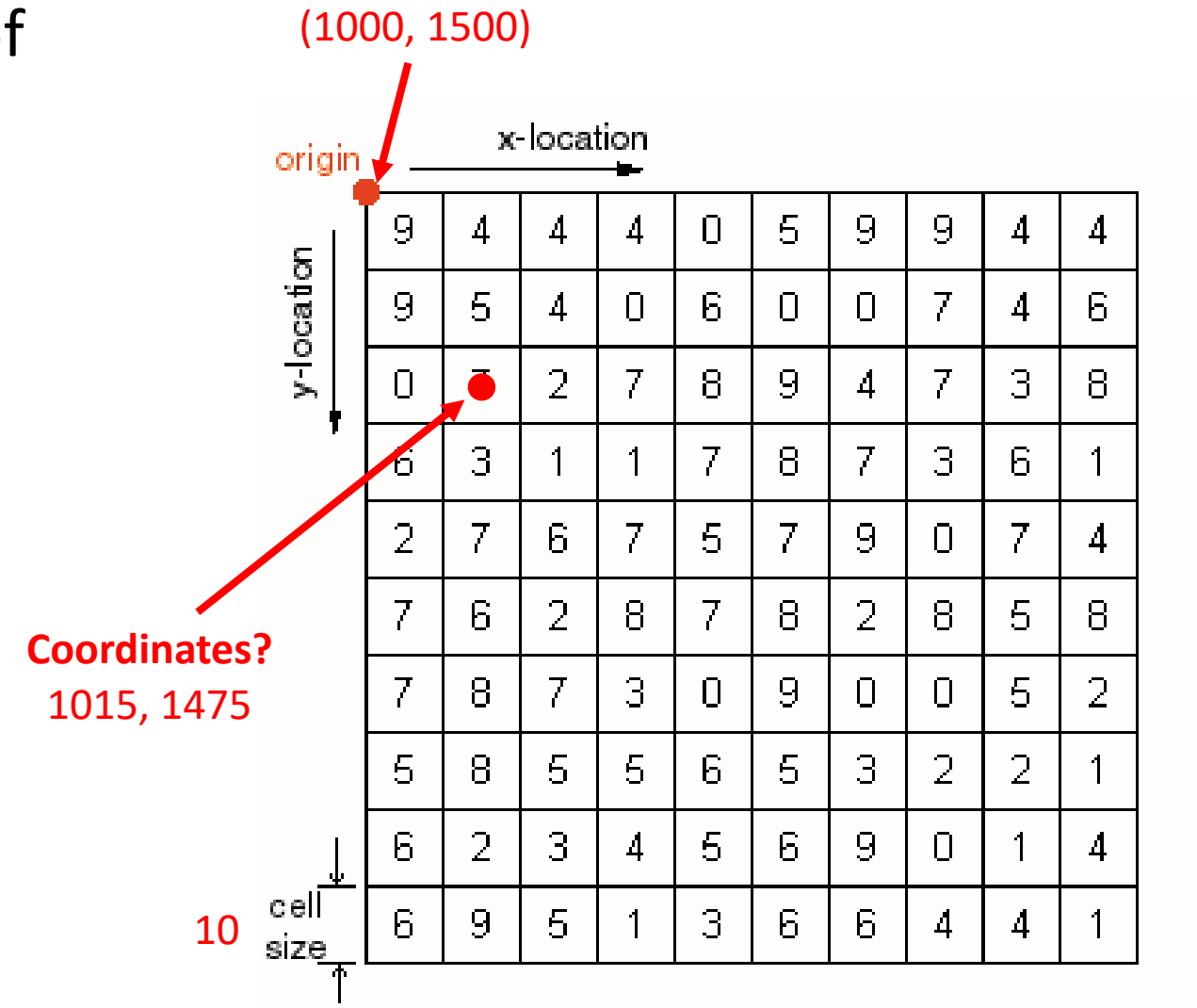
Spatial extent of the raster

Georeference of raster data

- A raster only stores the coordinates of the top-left corner – different from the vector model
- Coordinates of all other cells can be calculated from the coordinates at the top-left corner.

$$X = X_{top-left} + column\# * cell\_size + cell\_size$$

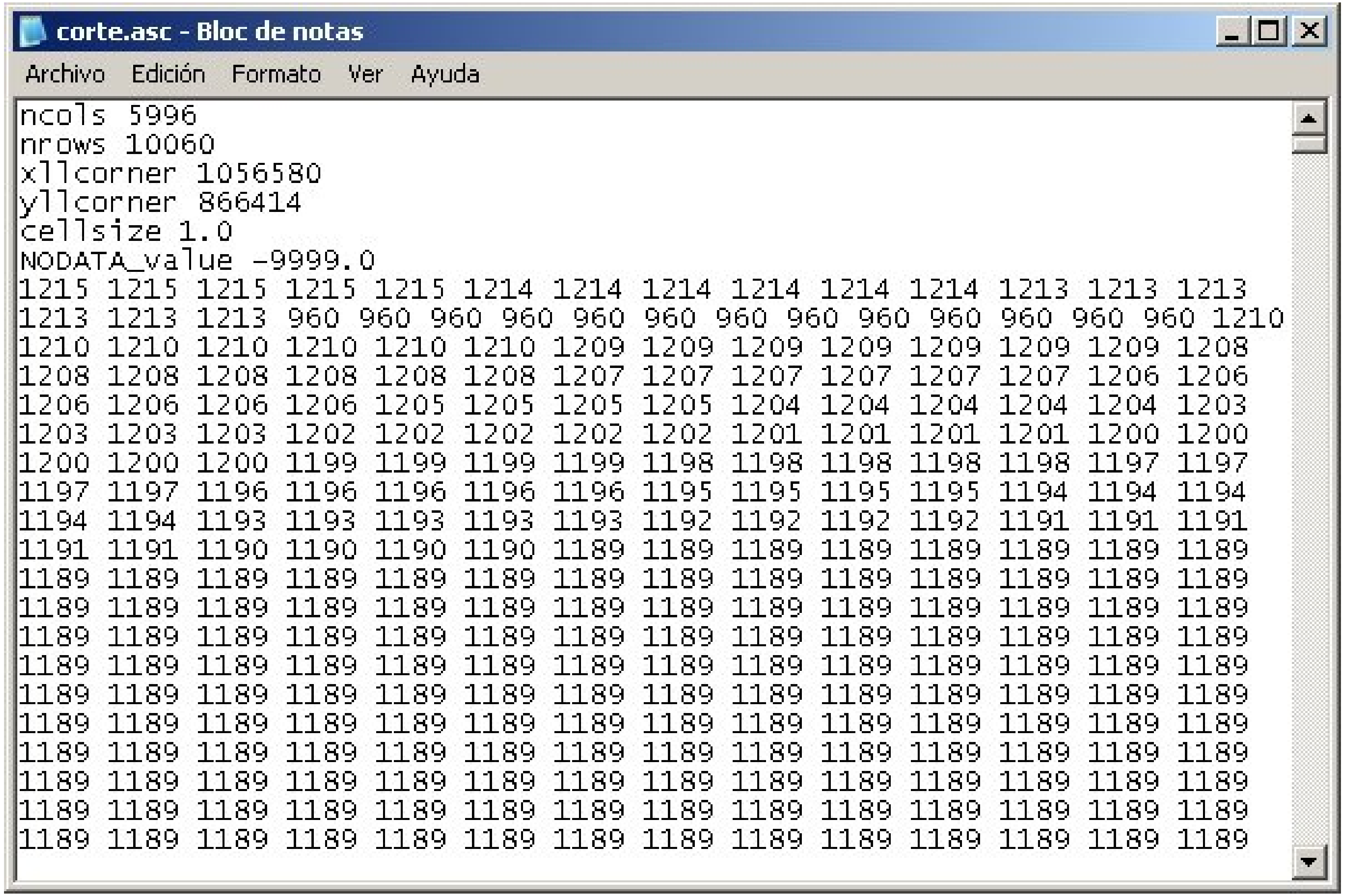
$$Y = Y_{top-left} - row\# * cell\_size + cell\_size$$



Raster data format

- A raster is essentially a spatially referenced 2D matrix.
- Additional information stored in a raster dataset:
  - Coordinate system, e.g. GCS, UTM, State Plane….
  - Coordinates of the top-left corner, e.g. [135000, 24620]
  - Cell size (the length of cell side), e.g. 30, 100, 1000
  - Horizontal unit, e.g. feet, meter, km, degree…

Raster in the ASCII Format


corte.asc - Bloc de notas
Archivo Edición Formato Ver Ayuda

```
ncols 5996
nrows 10060
xllcorner 1056580
yllcorner 866414
cellsize 1.0
NODATA_value -9999.0
1215 1215 1215 1215 1215 1214 1214 1214 1214 1214 1214 1213 1213 1213
1213 1213 1213 960 960 960 960 960 960 960 960 960 960 960 960 960 1210
1210 1210 1210 1210 1210 1210 1209 1209 1209 1209 1209 1209 1209 1208
1208 1208 1208 1208 1208 1208 1207 1207 1207 1207 1207 1207 1206 1206
1206 1206 1206 1206 1205 1205 1205 1205 1204 1204 1204 1204 1204 1203
1203 1203 1203 1202 1202 1202 1202 1202 1201 1201 1201 1201 1200 1200
1200 1200 1200 1199 1199 1199 1199 1198 1198 1198 1198 1198 1197 1197
1197 1197 1196 1196 1196 1196 1196 1195 1195 1195 1195 1194 1194 1194
1194 1194 1193 1193 1193 1193 1193 1192 1192 1192 1192 1191 1191 1191
1191 1191 1190 1190 1190 1190 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189 1189
```

Still needs a coordinate system!

Readings

- De Smith, Michael John, Michael F. Goodchild, and Paul Longley. *Geospatial analysis: a comprehensive guide to principles, techniques and software tools.*
  - Chapter 2.1: Basic primitives
    - 2.1.1 – 2.1.6

Online version: https://www.spatialanalysisonline.com/HTML/index.html