



GEODATABASES AND QUERIES



Databases

Why do we care about databases?



- Integrated sets of data
- Focused on a particular area and subject
- Form the basis of GIS Analysis
- Must be well-structured in order for us to best access the information stored in them

How are databases managed

- Database Management System (DBMS)
 - ▣ system or software program(s) that enables you to store, modify and extract information from a database

Relational Database

- Supports the representation of data as a set of *tables* that are *related* to each other
- Each table = a list of records containing attributes about features
- Tied together by an ‘Attribute Key’ or an attribute field common to both tables

Table data structure

- **Record:** a row in a database; represents one feature (a.k.a. “tuple” OR “row”)
- **Attribute:** a column in a database; contains attribute values (a.k.a. “field”, “item”, OR “variable”)

Attribute
or Item

	Name	FIPS	Pop90	Area	PopDn
	Whatcom	53073	128	2170	59
	Skagit	53057	80	1765	45
Record	Clallam	53009	56	1779	32
	Snohomish	53061	466	2102	222
	Island	53029	60	231	261
	Jefferson	53031	20	1773	11
	Kitsap	53035	190	391	485

Table data structure

Forests

Forest Name	Forest-ID	Location	Size
Nantahala	1	N. Carolina	184,447
Cherokee	2	N. Carolina	92,271

Trails

Trail Name	Forest-ID
Bryson's Knob	1
Slickrock Falls	2
North Fork	1
Cade's Cove	1
Cade's Cove	2
Appalachian	1
Appalachian	2

Recreational features

Feature	Description	Activity1	Activity2
Wfall	Waterfall	Photography	Swimming
Ogrth	Old-Growth Forest	Photography	Hiking
Vista	Scenic Overlook	Photography	Viewing
Wlife	Wildlife Viewing	Photography	Birding
Cmp	Camping	Camping	-

Characteristics

Trail Name	Feature	Difficulty
Bryson's Knob	Vista	E,M
Bryson's Knob	Ogrth	E,M
Slickrock Falls	Ogrth	M
Slickrock Falls	Wfall	M
North Fork	-	M
Cade's Cove	Ogrth	E
Cade's Cove	Wlife	E
Appalachian	Wfall	M,D
Appalachian	Ogrth	M,D
Appalachian	Vista	M,D
Appalachian	Wlife	M,D
Appalachian	Cmp	M,D

Which attribute is the 'key'?

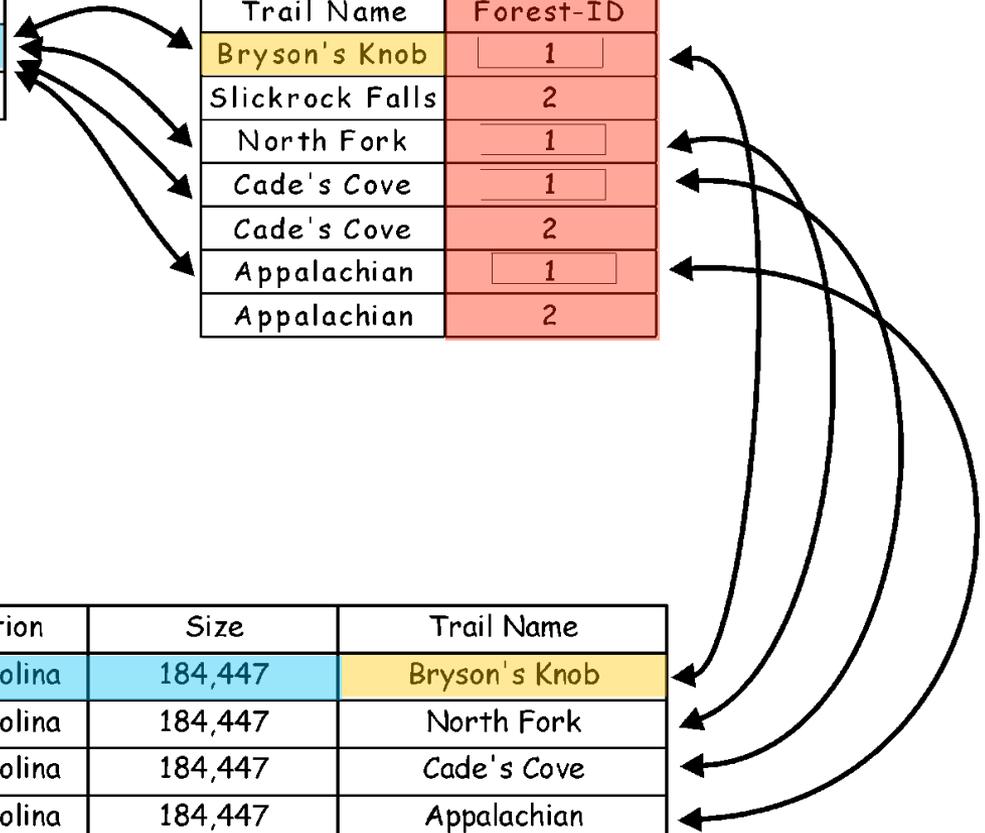
Forests

Forest Name	Forest-ID	Location	Size
Nantahala	1	N. Carolina	184,447
Cherokee	2	N. Carolina	92,271

Trails

Trail Name	Forest-ID
Bryson's Knob	1
Slickrock Falls	2
North Fork	1
Cade's Cove	1
Cade's Cove	2
Appalachian	1
Appalachian	2

Forest Name	Forest-ID	Location	Size	Trail Name
Nantahala	1	N. Carolina	184,447	Bryson's Knob
Nantahala	1	N. Carolina	184,447	North Fork
Nantahala	1	N. Carolina	184,447	Cade's Cove
Nantahala	1	N. Carolina	184,447	Appalachian
Cherokee	2	N. Carolina	92,271	Slickrock Falls
Cherokee	2	N. Carolina	92,271	Cade's Cove
Cherokee	2	N. Carolina	92,271	Appalachian



How are tables 'related'?

Forests

Forest Name	Forest-ID	Location	Size
Nantahala	1	N. Carolina	184,447
Cherokee	2	N. Carolina	92,271

Trails

Trail Name	Forest-ID
Bryson's Knob	1
Slickrock Falls	2
North Fork	1
Cade's Cove	1
Cade's Cove	2
Appalachian	1
Appalachian	2

Characteristics

Trail Name	Feature	Difficulty
Bryson's Knob	Vista	E,M
Bryson's Knob	Ogrth	E,M
Slickrock Falls	Ogrth	M
Slickrock Falls	Wfall	M
North Fork	-	M
Cade's Cove	Ogrth	E
Cade's Cove	Wlife	E
Appalachian	Wfall	M,D
Appalachian	Ogrth	M,D
Appalachian	Vista	M,D
Appalachian	Wlife	M,D
Appalachian	Cmp	M,D

Recreational features

Feature	Description	Activity1	Activity2
Wfall	Waterfall	Photography	Swimming
Ogrth	Old-Growth Forest	Photography	Hiking
Vista	Scenic Overlook	Photography	Viewing
Wlife	Wildlife Viewing	Photography	Birding
Cmp	Camping	Camping	-

Object (ODBMS) & Object-Relational (ORDBMS)

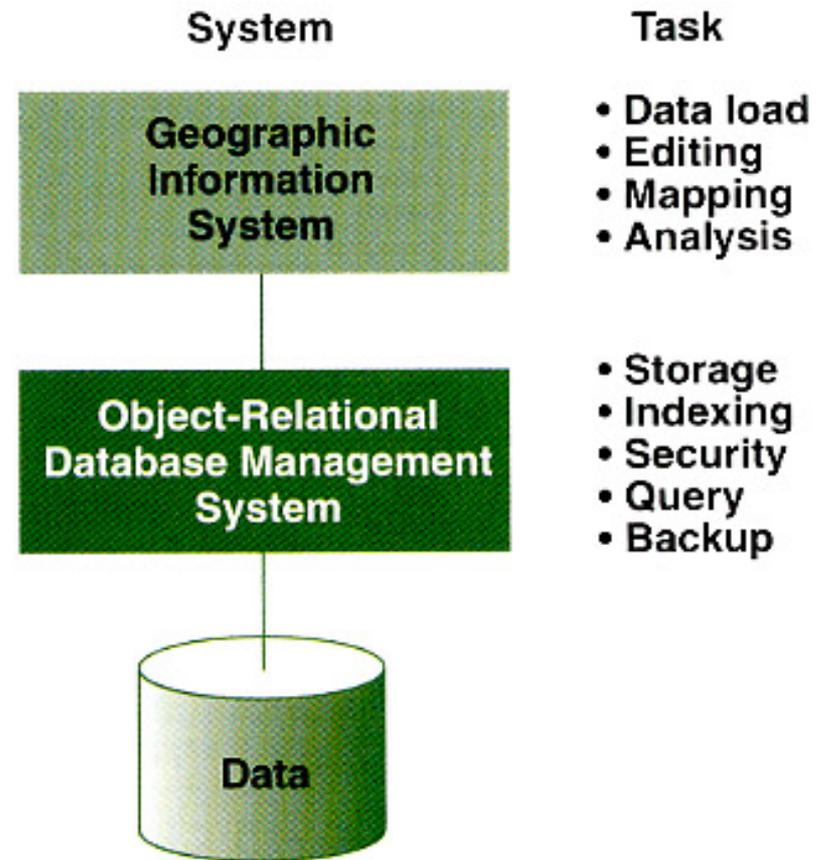
- ODBMS supports the representation of data as objects having attributes, methods, & behaviors
 - ▣ ArcGIS geodatabases
- Designed to address the weaknesses of RDBMS such as
 - ▣ Geometry and attribute data are stored in separate databases
 - ▣ Poor performance for many types of geographic query
- Geographic extensions to standard RDBMS can provide similar functionality (ORDBMS)



Geographic (spatial) databases

What is a geographic database?

- Contains one or more tables with a geographic component (a “shape” attribute)
- Common example: an ESRI Geodatabase



ESRI Geodatabase

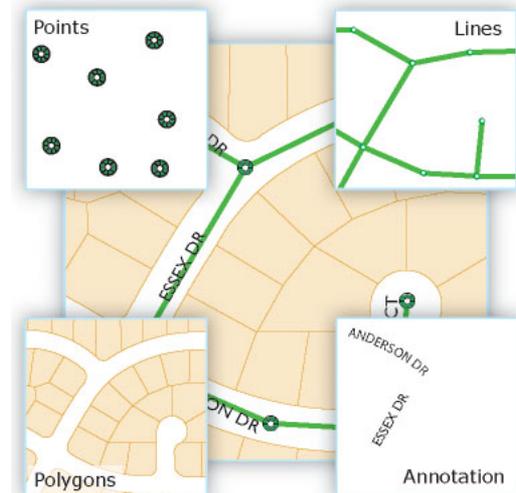
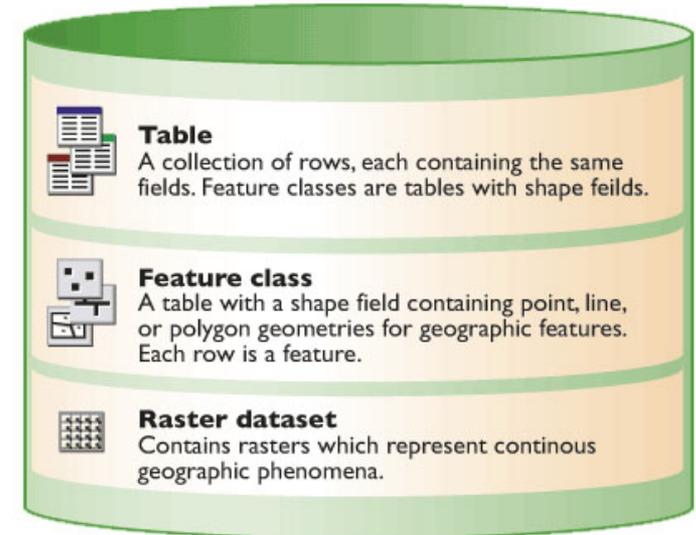
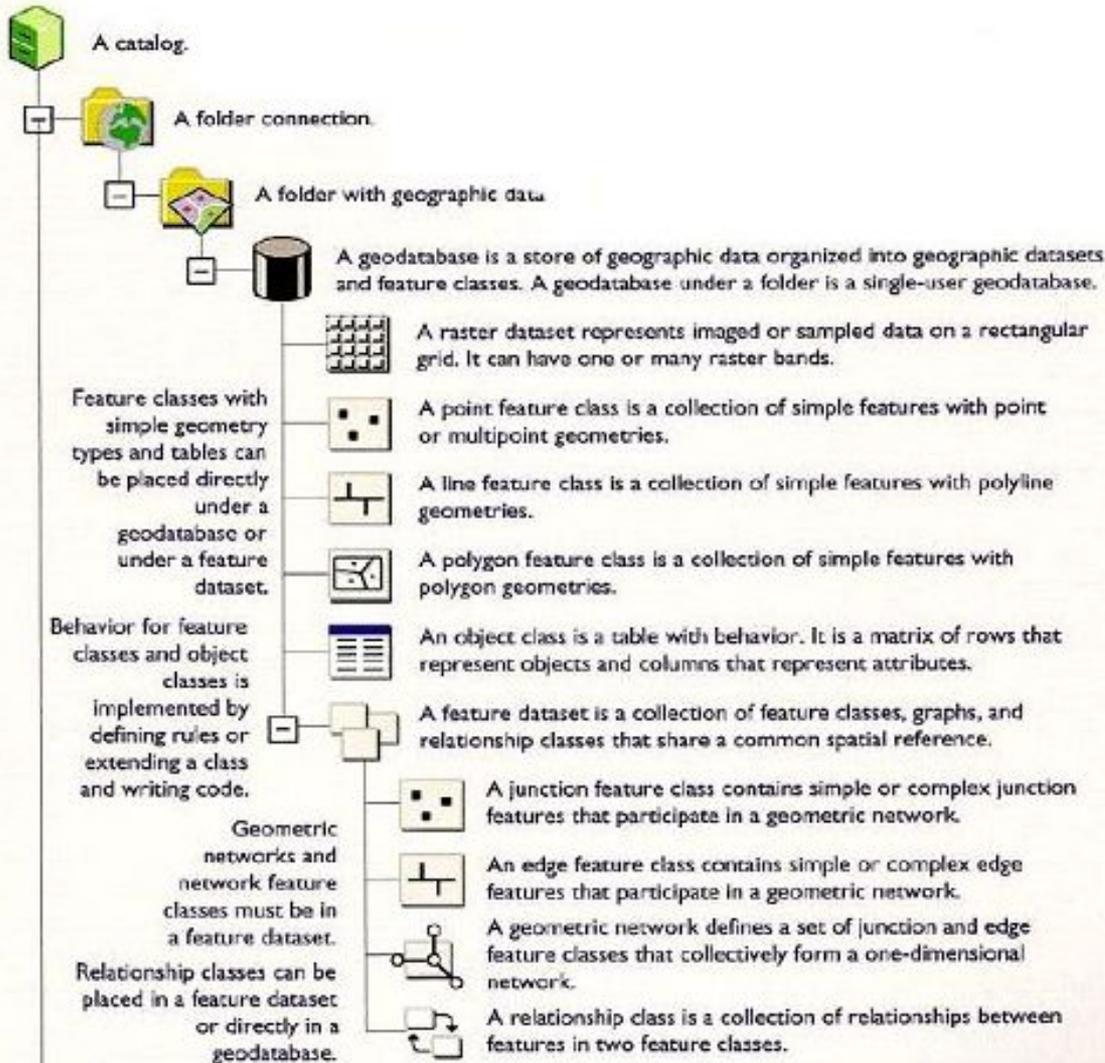
□ Features have:

- Shape
- Attribute
- Spatial reference
- Relationships

□ Features can be:

- Constrained by domains
- Validated by rules
- Modeled with complex behavior

ESRI Geodatabase

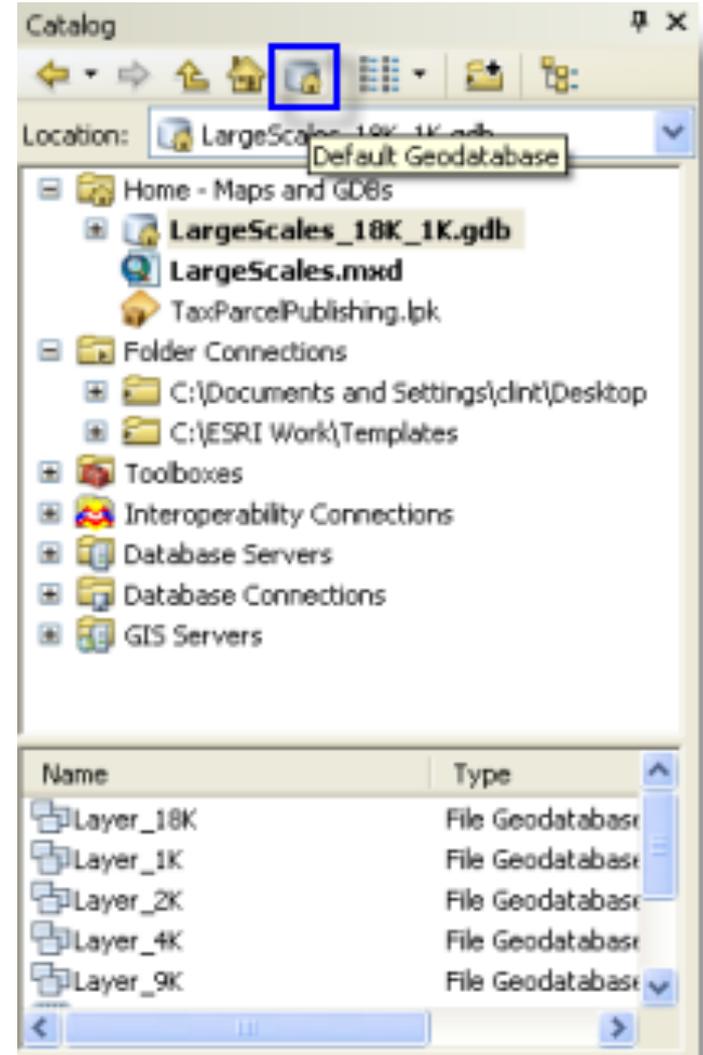


Types of Geodatabases

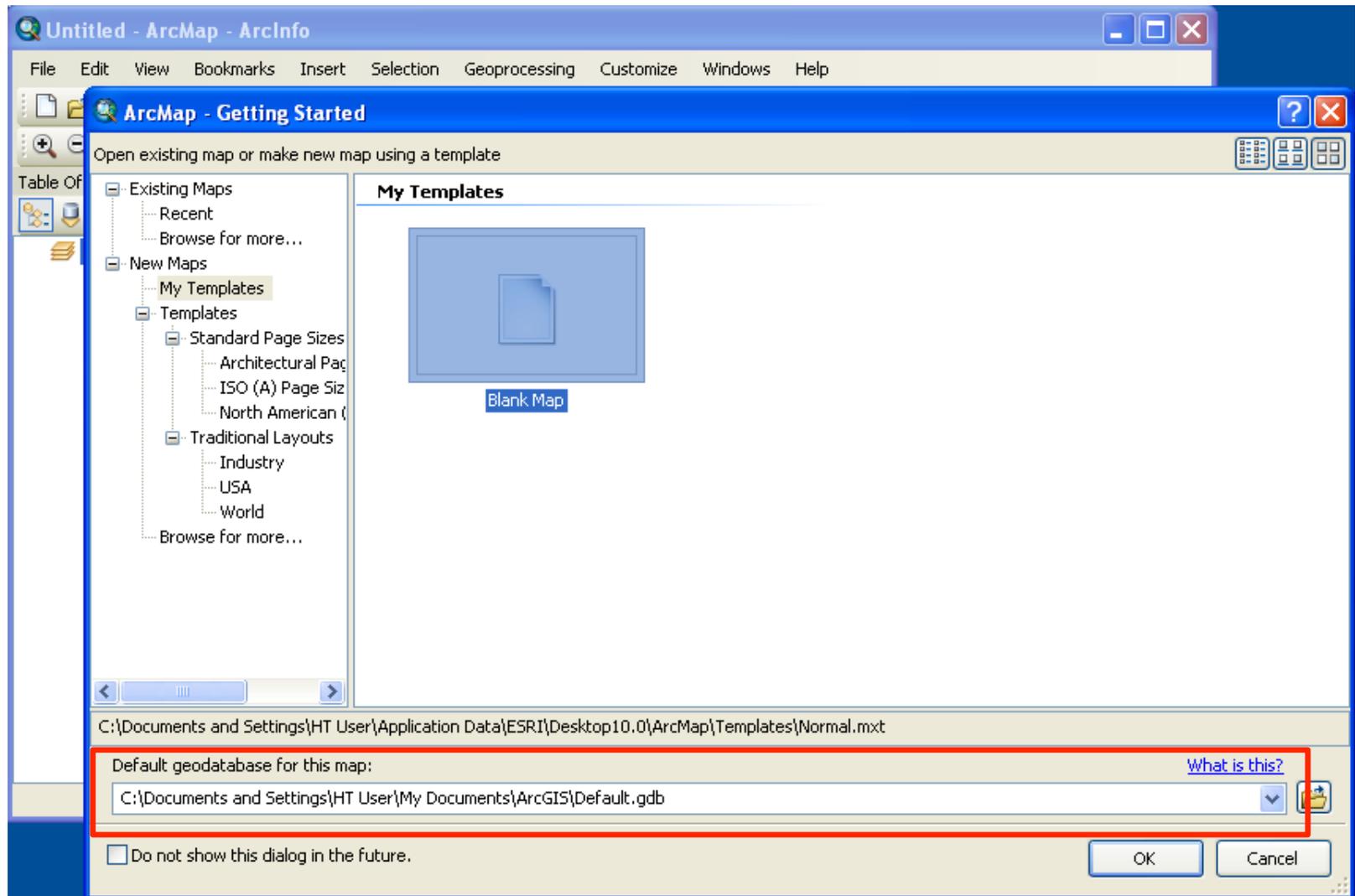
- Personal
 - ▣ Single-user
 - ▣ Microsoft Access
 - ▣ Up to 2GB storage
- File
 - ▣ Single-user
 - ▣ File folder structure
 - ▣ Up to 1TB storage
- ArcSDE
 - ▣ Multi-user, supports versioning
 - ▣ Oracle, Microsoft SQL server, IBM DB2, IBM Informix, Postgre SQL
 - ▣ Storage limit based on DBMS type

“Default” Geodatabase

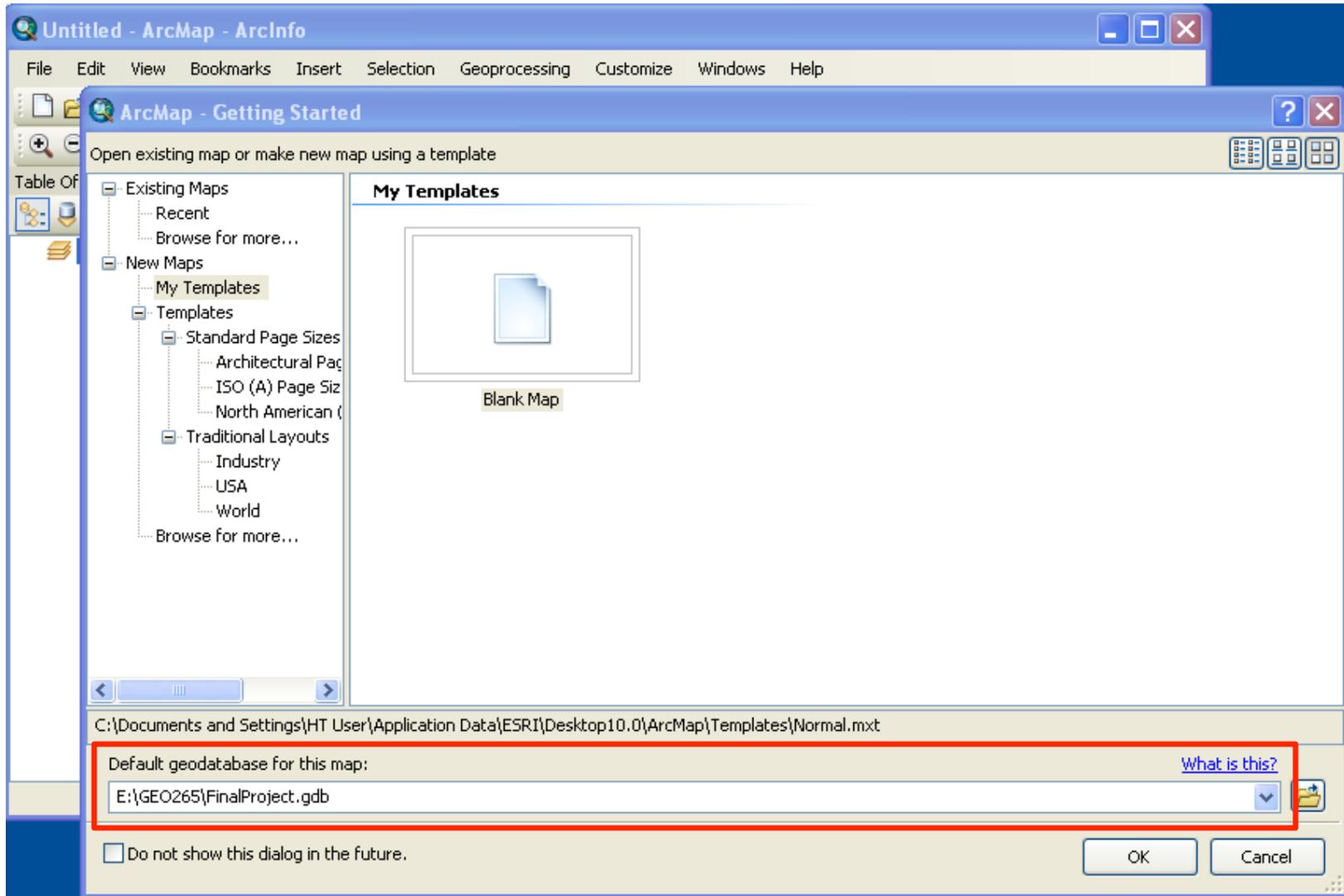
- A setting in ArcGIS
- Helps with data management
- Allows you to:
 - ▣ Store all datasets in one convenient location, or
 - ▣ Set a database for each project



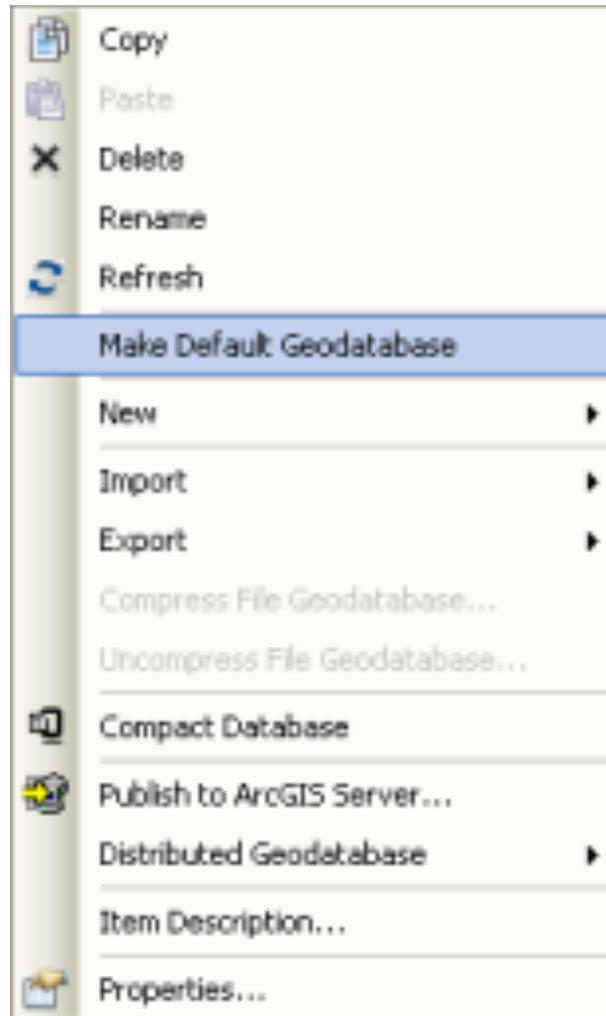
Setting the Default Geodatabase - ArcMap



Change location of Default Geodatabase



Setting the Default Geodatabase – ArcCatalog





Querying GIS data

Getting information from GIS data

- Query is a question or inquiry about a feature(s) shown on a map (and therefore, in a GIS database)
- Output is a selected set of records
- Two approaches
 - ▣ Attribute queries
 - ▣ Spatial queries

Attribute query

- Selects features based on *non-spatial* information in the attribute table
 - ▣ For example: Which cities in the US have more than 100,000 people
- ‘Select by Attributes’ tool in ArcMap
- Uses SQL (structured query language) used to write the query

Simple selection:

records with Area > 20.0

ID	Area	Landuse	Municip
1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
5	22.1	Urban	City
6	30.2	Farm	Township
7	4.4	Urban	County

AND selection:

records with (Landuse = Urban) and
(Municip = City)

ID	Area	Landuse	Municip
1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
5	22.1	Urban	City
6	30.2	Farm	Township
7	4.4	Urban	County

OR selection:

records with (Area > 20.0)
OR (Municip = City)

ID	Area	Landuse	Municip
1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
5	22.1	Urban	City
6	30.2	Farm	Township
7	4.4	Urban	County

NOT selection:

records with
Landuse NOT Urban

ID	Area	Landuse	Municip
1	10.5	Urban	City
2	330.3	Farm	County
3	2.4	Suburban	Township
4	96.0	Suburban	County
5	22.1	Urban	City
6	30.2	Farm	Township
7	4.4	Urban	County

Query Exercise

Row ID	Country	Population (millions)	Energy Use per Capita (brls oil)	Infant Mortality (per 1000)	Life Expectancy (years)	Car Theft (%)
1	Australia	19.9	5,668	4	79.2	2.2
2	Britain	59.3	5,945	5	77.5	2.6
3	Finland	5.2	6,456	4	78.0	0.5
4	France	59.7	4,350	4	79.2	1.8
5	Japan	127.2	4,071	3	81.6	0.1
6	Netherlands	16.2	5,993	5	78.3	0.5
7	Norway	4.6	6,019	4	78.9	1.5
8	South Africa	45.3	3,703	52	46.5	2.4
9	Spain	41.1	2,945	5	78.3	0.5
10	U.S.A.	291.0	8,066	7	77.3	0.5

Identify the rows in the table that meet the following criteria:

- Per capita energy use >6000 AND population <20 million
- Infant mortality <7 AND life expectancy >79.0
- Per capita energy use <4000 OR [(population > 40 million) AND (car theft <1)]
- [(Per capita energy use <4000) OR (population > 40 million)] AND (car theft <1)
- NOT (population $>40,000,000$)
- Population $<20,000,000$ AND NOT (car theft >1.5)

Definition query

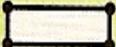
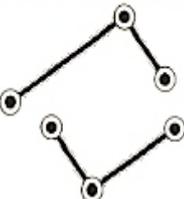
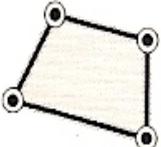
- Also an attribute query, but ‘defines’ what is visible on the map
- Does not select or highlight features on the map or in the attribute table

Spatial query

- Selects features based on *location* or *spatial relationship* between data layers
 - ▣ Which Oregon counties have an interstate highway passing through them?
- ‘Select Layer by Location’ tool in ArcMap
 - ▣ Target layer – base layer, contains the features you want selected
 - ▣ Filter layer – comparison layer
- What kind of spatial relationships are possible between points, lines and polygons?

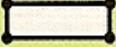
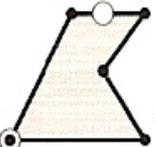
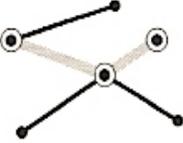
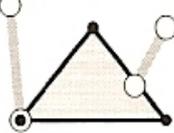
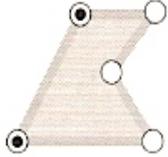
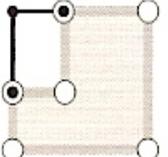
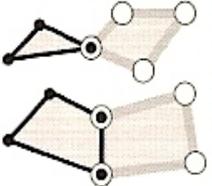
Equals

Does the base equal the comparison geometry?

		Base Geometry		
				
Comparison Geometry			No equals relationship possible	No equals relationship possible
		No equals relationship possible		No equals relationship possible
		No equals relationship possible	No equals relationship possible	

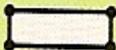
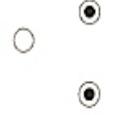
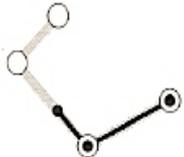
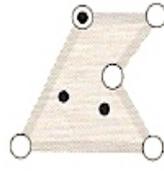
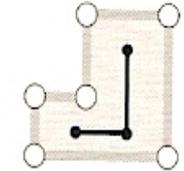
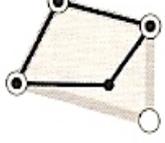
Touches

Does the base geometry touch the comparison geometry?

		Base Geometry		
				
Comparison Geometry		No touch relationship possible		
				
				

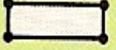
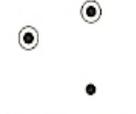
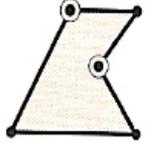
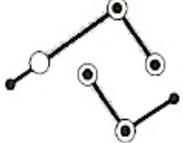
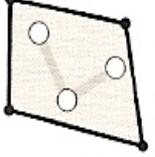
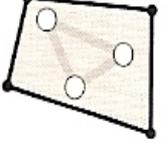
Within

Is the base within the comparison geometry?

		Base Geometry		
				
Comparison Geometry			No within relationship possible	No within relationship possible
				No within relationship possible
				

Contains

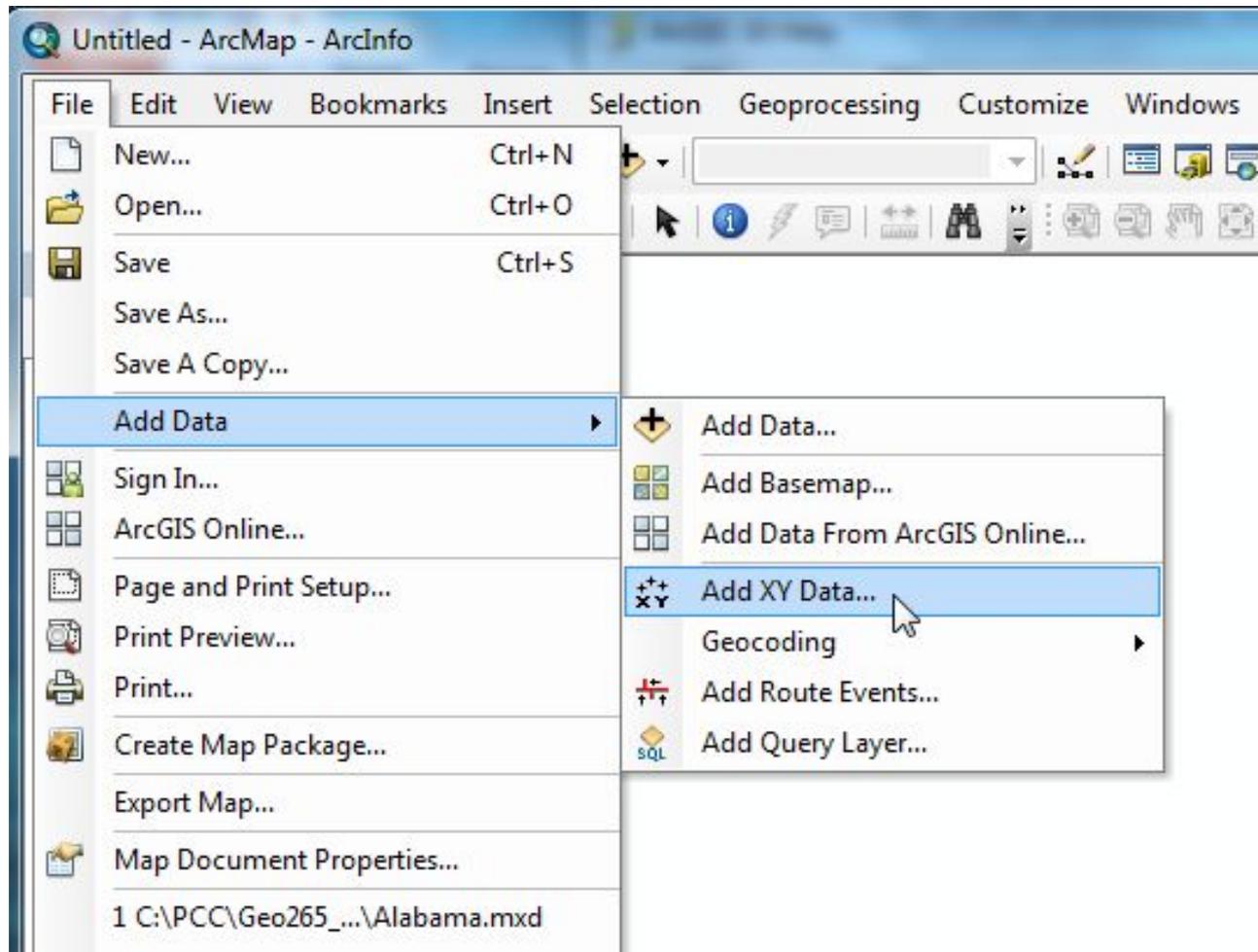
Does the base geometry contain the comparison geometry?

		Base Geometry		
				
Comparison Geometry				
		No containment relationship possible		
		No containment relationship possible	No containment relationship possible	



Adding data to a map

“Add XY Data” tool



Bakery list – Excel file

The screenshot shows a Microsoft Excel window titled 'pdx_bakeries.xlsx'. The ribbon is set to 'View'. The spreadsheet contains a table with the following data:

	A	B	C	D	E	F	G
1	ID	NAME	YCOORD(LAT)	XCOORD(LON)	BESTDESSERT	PARKING	
2	1	St. Honore	45.535649	-122.699627	Millefeuille	Ok	
3	2	Cupcake Jones	45.525343	-122.681378	Red Velvet Cupcake	Bad	
4	3	Sweetpea	45.519506	-122.653481	Choc Chip Cookie	Good	
5	4	New Cascadia	45.510811	-122.659509	Caramel Rolls	Good	
6							
7							
8							
9							
10							

New Point Layer for Bakery Locations

