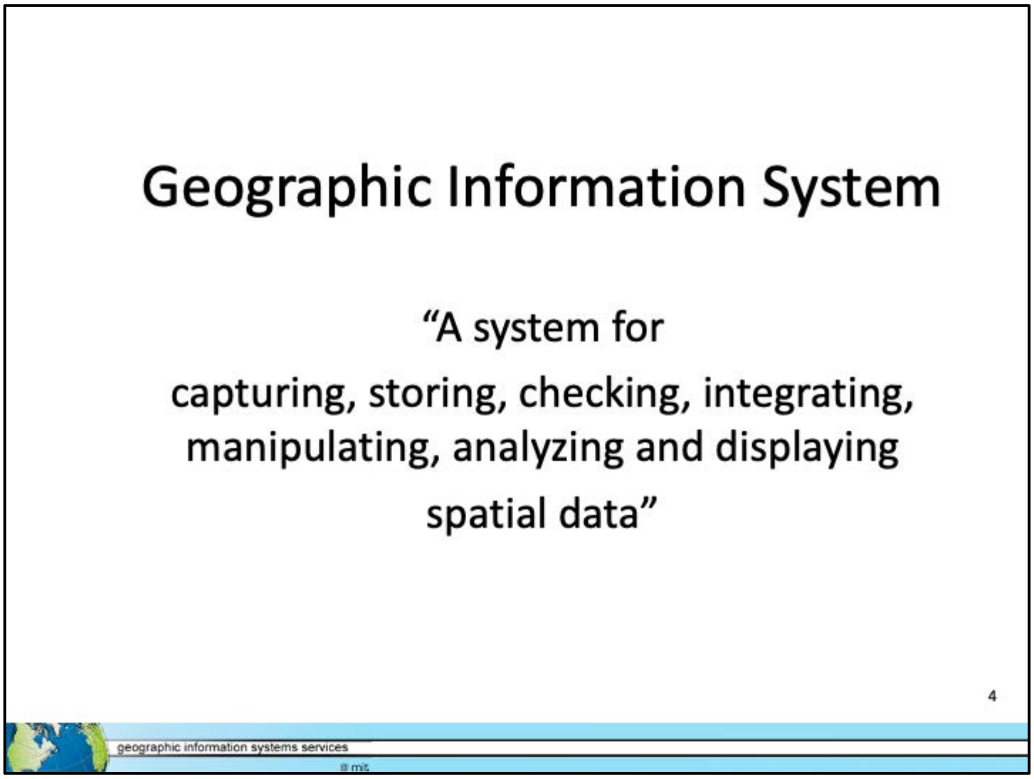

GIS Level 1: Introduction
to GIS & Mapping
1
Courtesy of US Air Force. Image is in the public domain.

Outline

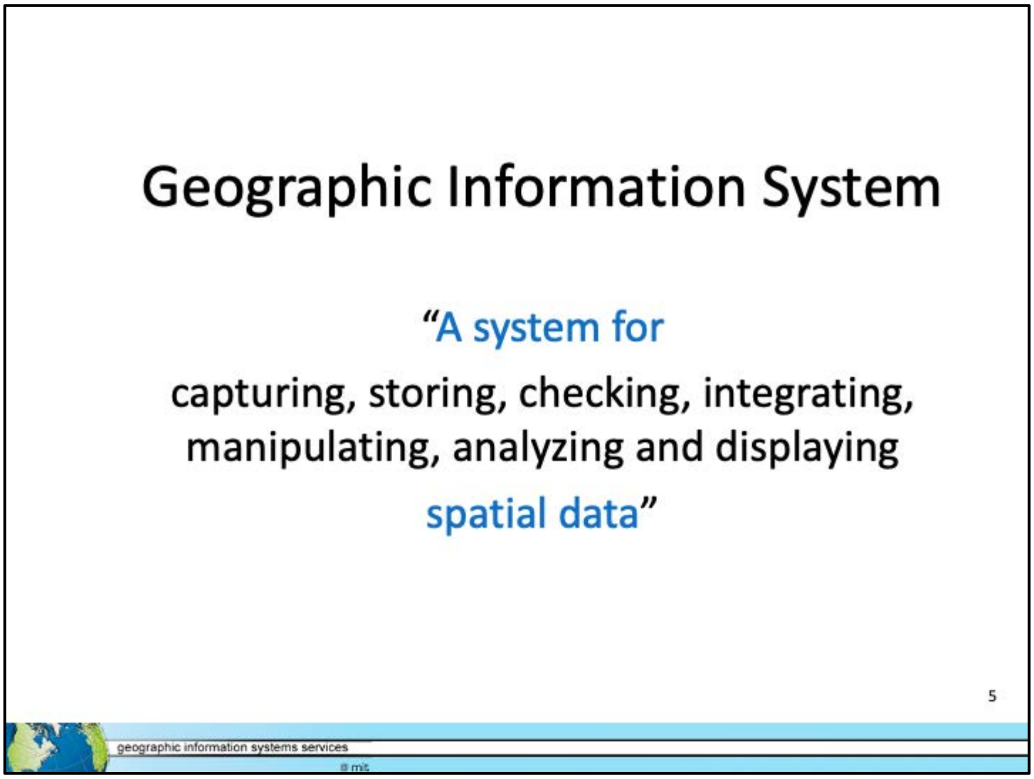
- Introduction – What is GIS?
- Software options
- Applications
- Understanding Maps & Data
  - Data Layers
  - Spatial Data Types
  - Characteristics of Spatial Data
  - Metadata
- Making Great Maps – Data Visualization Principles

INTRODUCTION

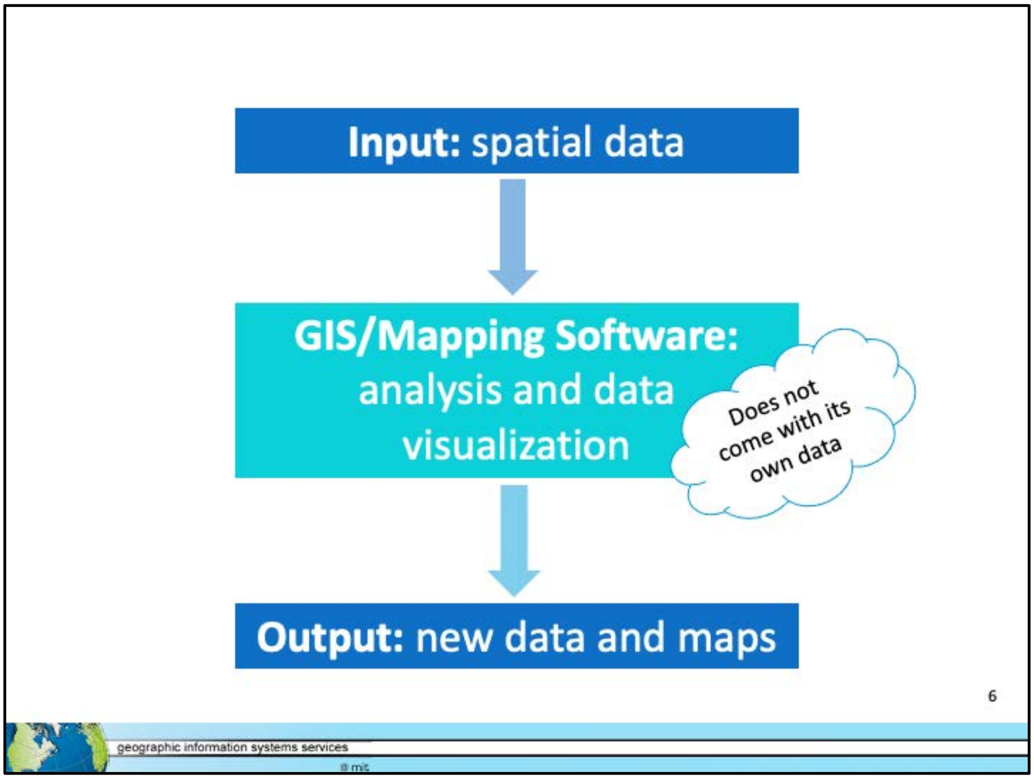


GIS stands for geographic information system but what does that really mean?

A GIS system can do a lot of things.



The most important thing to know is that a GIS is a system works with spatial data.



GIS software works upon the data you provide it.
Just like other software, such as Excel or Word, you need to input data and can then use the software to display and analyze that data.

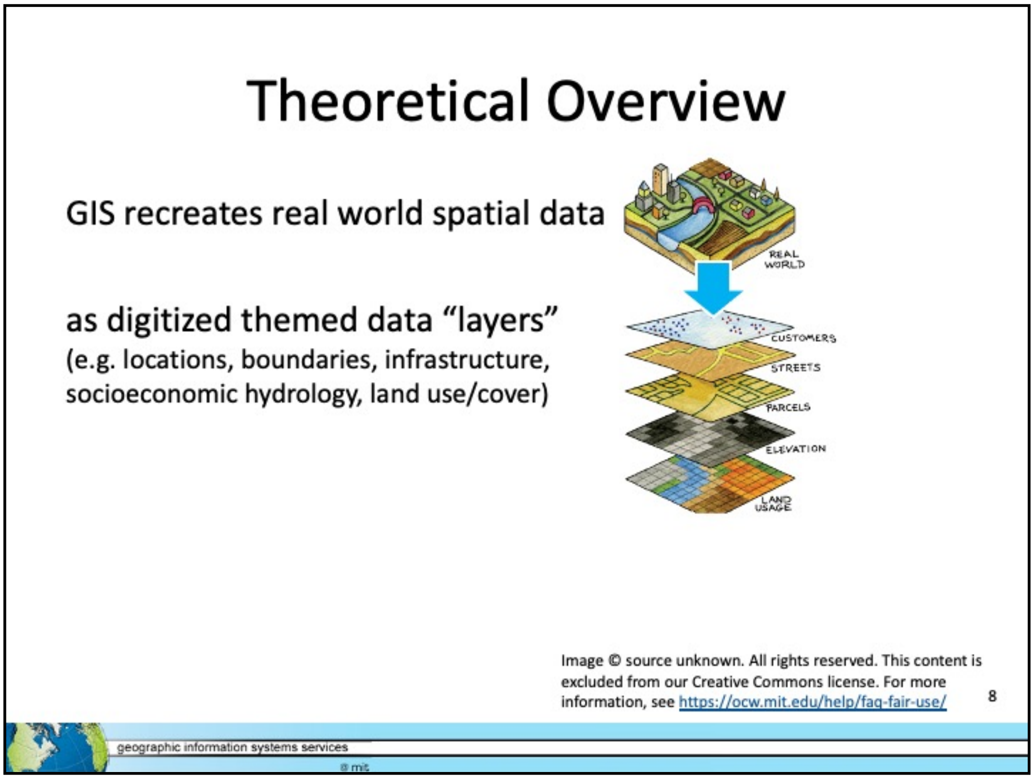
Theoretical Overview

GIS recreates real world spatial data

7

geographic information systems services @ mit

GIS software recreates or reproduces the real world as spatial data.



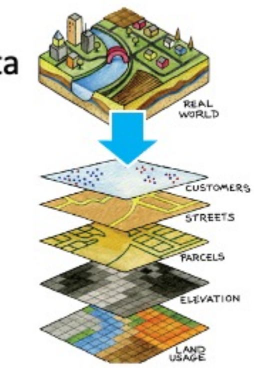
More specifically, the spatial data is broken down into themed data “layers” such as locations, boundaries of various kinds, socioeconomic variables, hydrology, and land use/cover.

Theoretical Overview

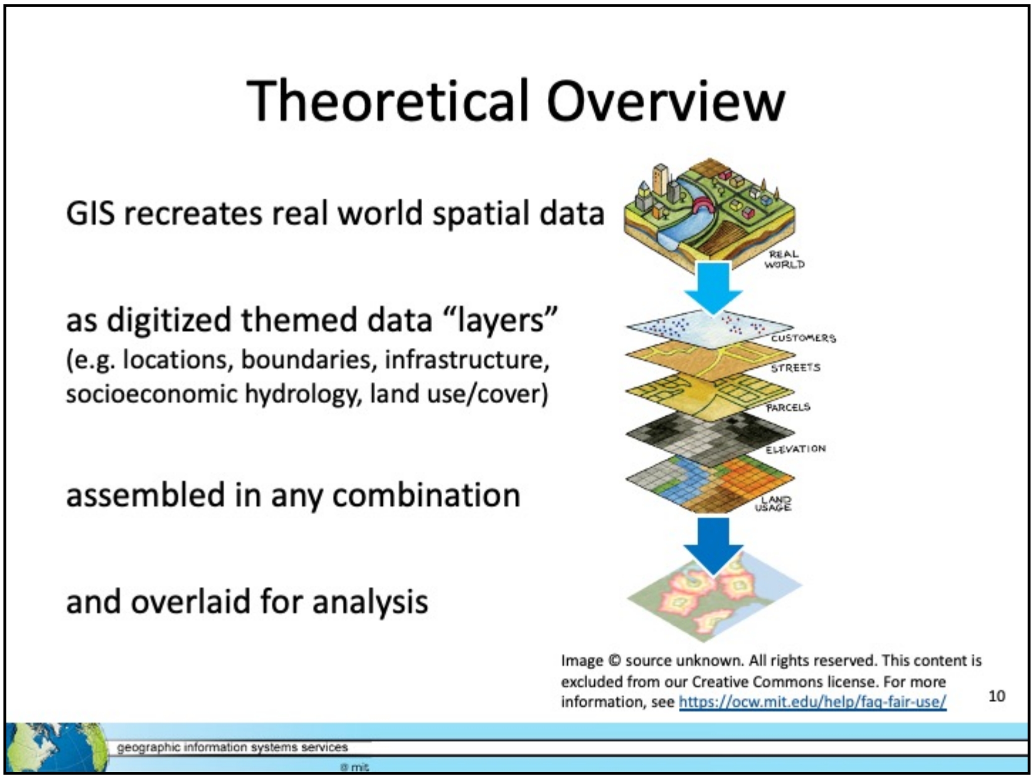
GIS recreates real world spatial data

as digitized themed data "layers"
(e.g. locations, boundaries, infrastructure, socioeconomic hydrology, land use/cover)

assembled in any combination



These data layers can be assembled in any combination you want, depending on what you want to highlight.



Power of GIS is in the ability to detect trends and make quantified decisions that would have otherwise been difficult just by hand or eye.

Once they have been overlaid in a particular combination, you can then conduct analyses.

SOFTWARE

11

geographic information systems services
@ mit

Types of GIS & Mapping Software

Type	Analysis Power	Example(s)
Geobrowser	**Weak** (mainly only to display data)	Google Maps, Google Earth, Apple Maps, Waze, etc.
Web-based	**Medium** (able to upload additional data, customize display, and perform basic analyses)	Carto, ArcGIS Online, Mapbox, Google MyMaps, etc.
Desktop	**Strong** (installed locally, provides full control of map creation, and perform advanced analyses)	ArcGIS Pro QGIS

The software you choose depends on what you will be doing with the data (displaying it vs. creating a new map vs. conducting analysis), the size of the data (large datasets require storage), and your audience (is it best to present the map on paper? Online? Do people need to interact with it?).

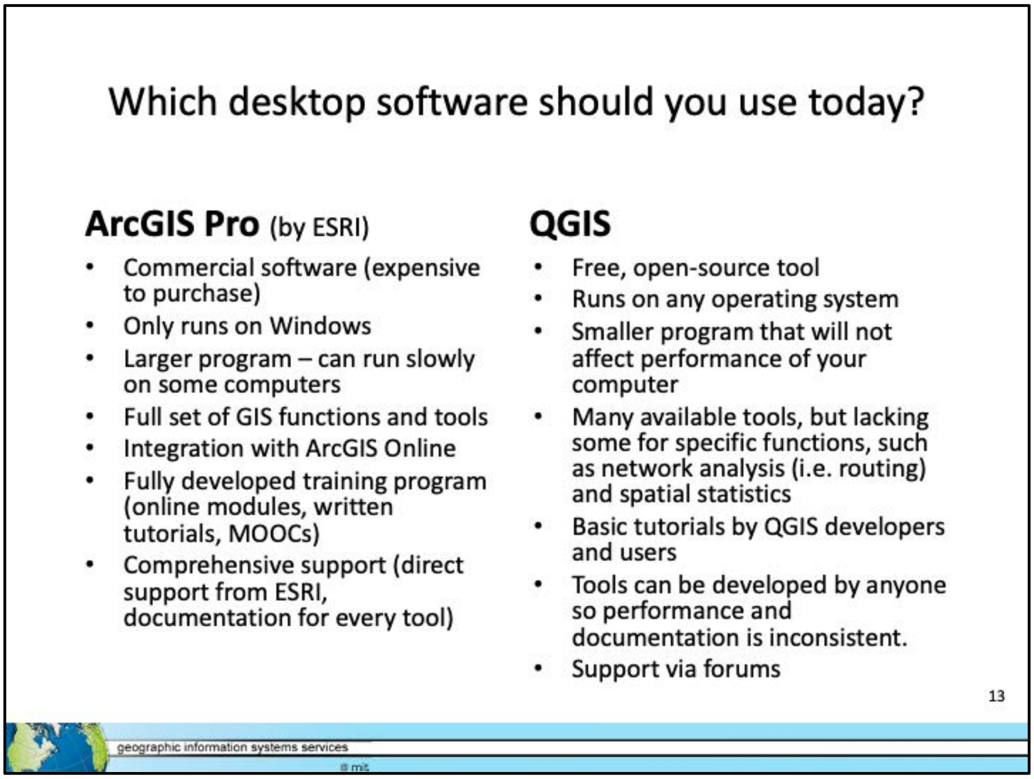
There are three types of software with very different capabilities.

1. Geobrowsers like Google Maps and Apple Maps are generally only useful for displaying data.

2. Web based tools like Carto, ArcGIS Online, and Mapbox, allow you to upload data, customize displays and perform basic analyses. Carto is licensed for MIT use and is free to use while you are at MIT but charges a fee otherwise. ArcGIS Online has both a free, public version and a paid version with additional features that MIT affiliates can access.

-Web-based software is the best way to add interactive elements. If you have programming skills you can use tools like Leaflet or OpenLayers to create online maps. Otherwise ArcGIS Online is easy to use and customize. Find out more about getting and account:
https://libguides.mit.edu/gis/webmap
Learn more about using ArcGIS Online: https://doc.arcgis.com/en/arcgis-online/create-maps/create-maps-and-apps.htm

3. Desktop software, which are installed locally, provide the fullest suite of GIS tools for basic and advanced analyses and map creation. You will get hands on experience for each type of GIS in this presentation series.



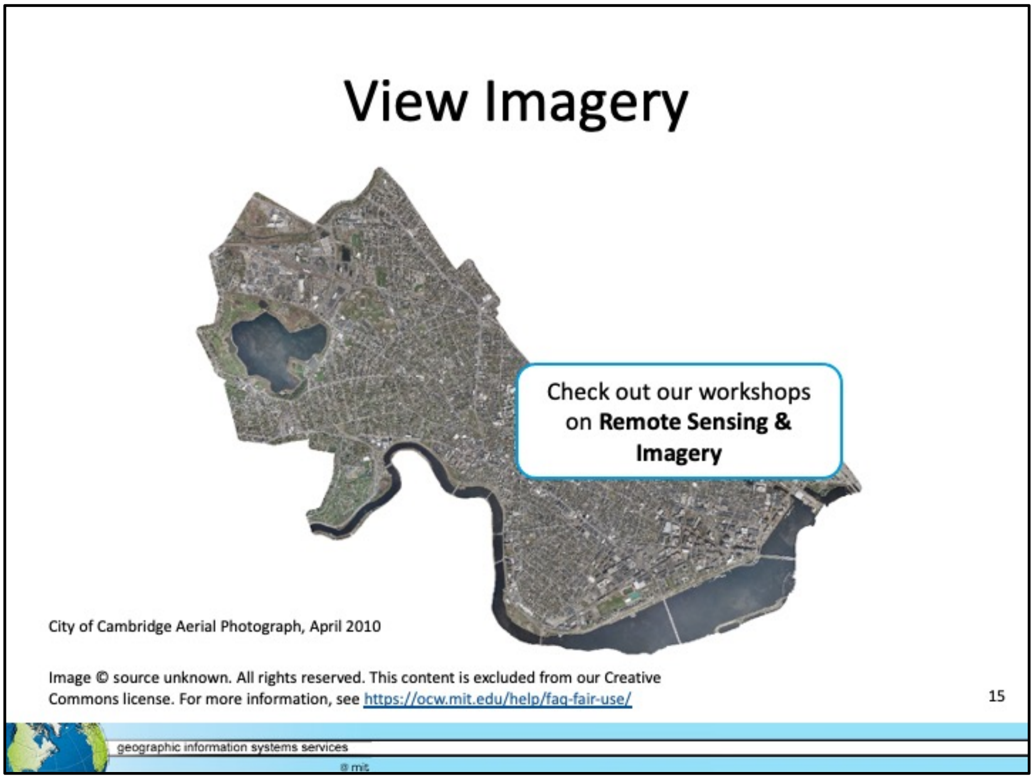
- Arcgis Pro is a more complete set of GIS tools but is license software and you may not have access to it beyond MIT. QGIS is free and open source but is a more limited set of GIS tools.
- Along with the comprehensive support from Esri for Arcgis Pro, there is complete documentation for all ArcGIS Pro tools. In QGIS tools are made by many people with varying resources. Because of this there performance and documentation across tools is inconsistent.
- Arcgis Pro only runs on Windows while Qgis runs on Windows, Apple, and Linux machines.
- Both interfaces are similar so after you learn one software, it is fairly easy to use the other
- For most uses, either software will work well so you may want to base your decision on your computer's resources, how frequently you will need to use it, what you want to do, and the industry you plan to work in (or are currently working in). ESRI products are frequently used in educational settings, municipal governments, and large businesses and organizations. Because of cost considerations, QGIS may be used in small businesses and non-profits.

GIS APPLICATIONS

14

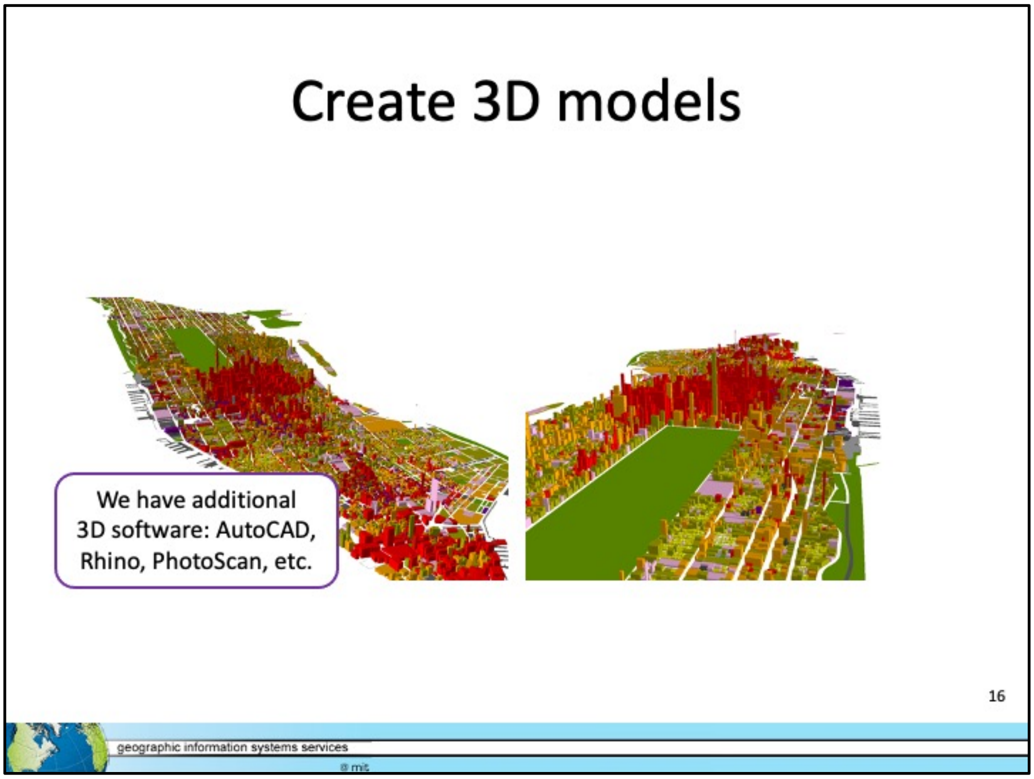
geographic information systems services
@ mit

Now let’s discuss some applications you can do with a GIS.



Many people use a GIS to look at satellite imagery or have it as a background to their maps.

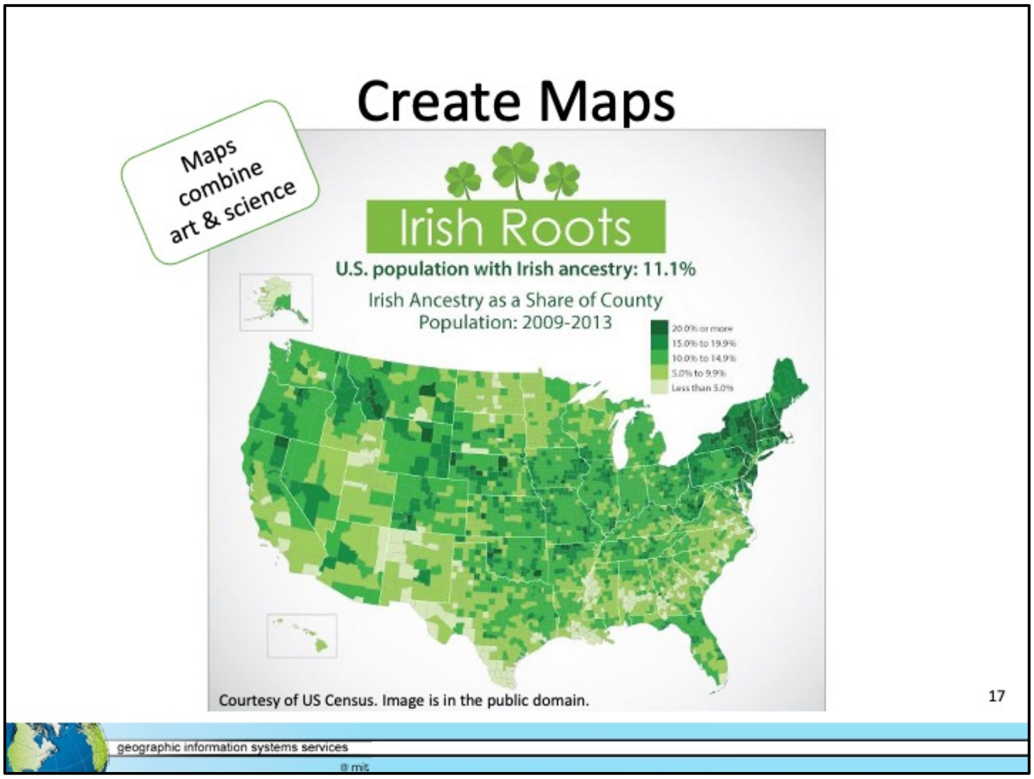
-materials from our remote sensing workshops are on our guide: libguides.mit.edu



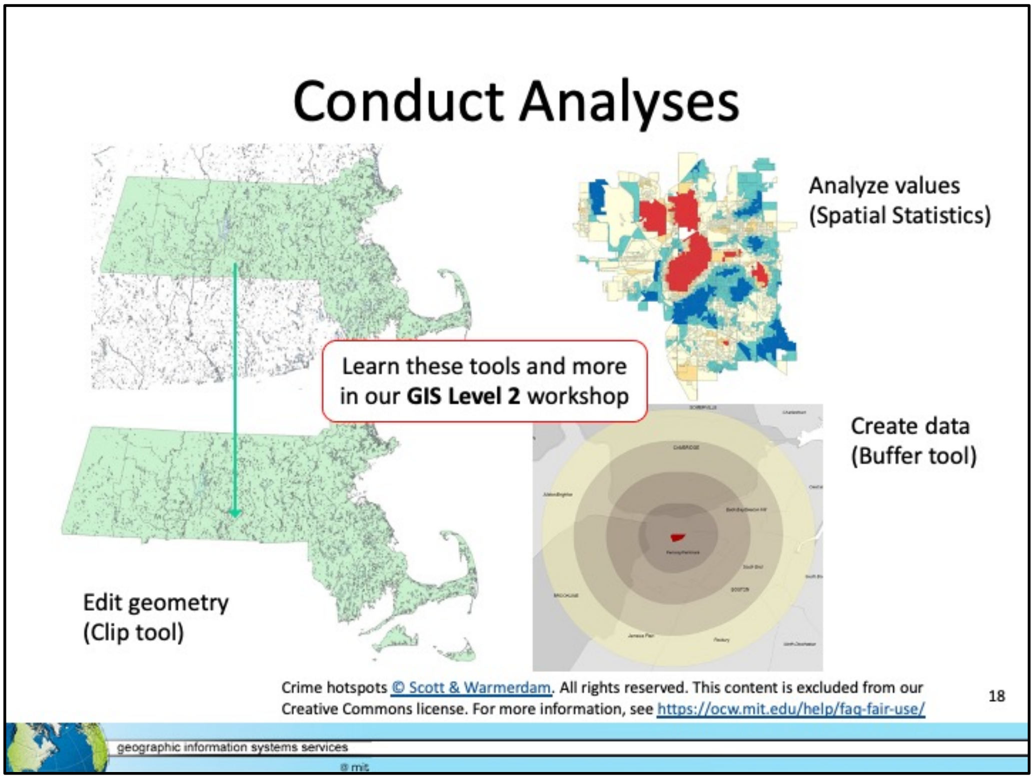
You can also make custom visualizations.

In this example, buildings in New York City have been color coded by land use and extruded vertically to represent height.

It is also possible to create animations over time and record videos flying through your map in a GIS.



Maps also are used to share information, where design plays an important (even persuasive) role.
See how the creator chose green for Irish ancestry, and a darker shade of green for higher percentage



GIS has many tools for acting upon datasets:
Some tools act on the geometry (clip), create new data (buffer), or analyze the data values (spatial statistics).

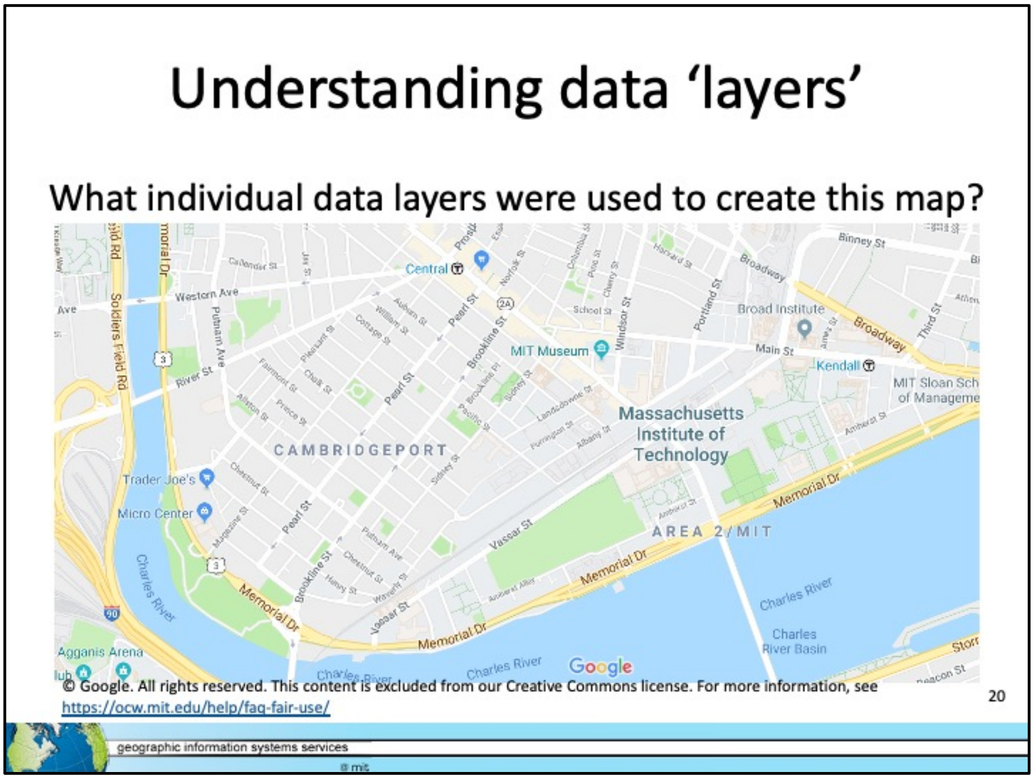
UNDERSTANDING MAPS & DATA

You may have been looking at geospatial data for a long time

19

geographic information systems services
@ mit

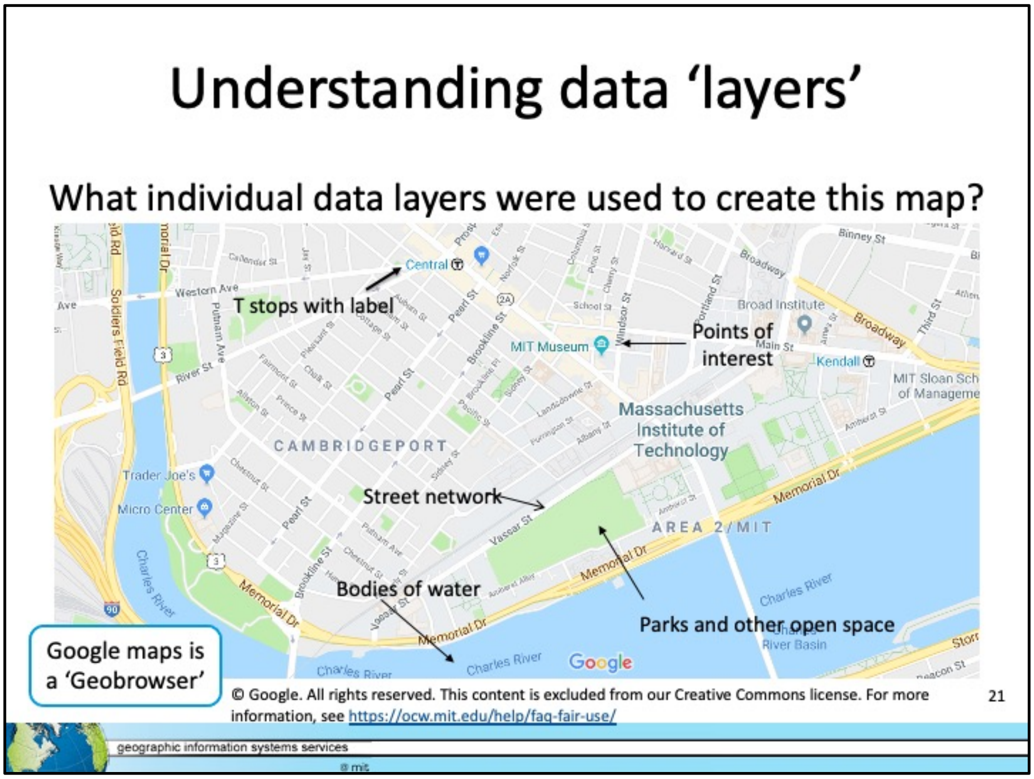
We will explore maps and data is this section of the presentation.



Google Maps is a type of GIS, really a web mapping tool or geobrowser, and its maps are created from multiple layers of spatial data. As shown previously, a layer is data about a specific set of similar features, such as the location of schools or bicycle paths.

Question: Take a moment to determine what distinct layers are in this map.

Presenter should move on after a few responses come in.

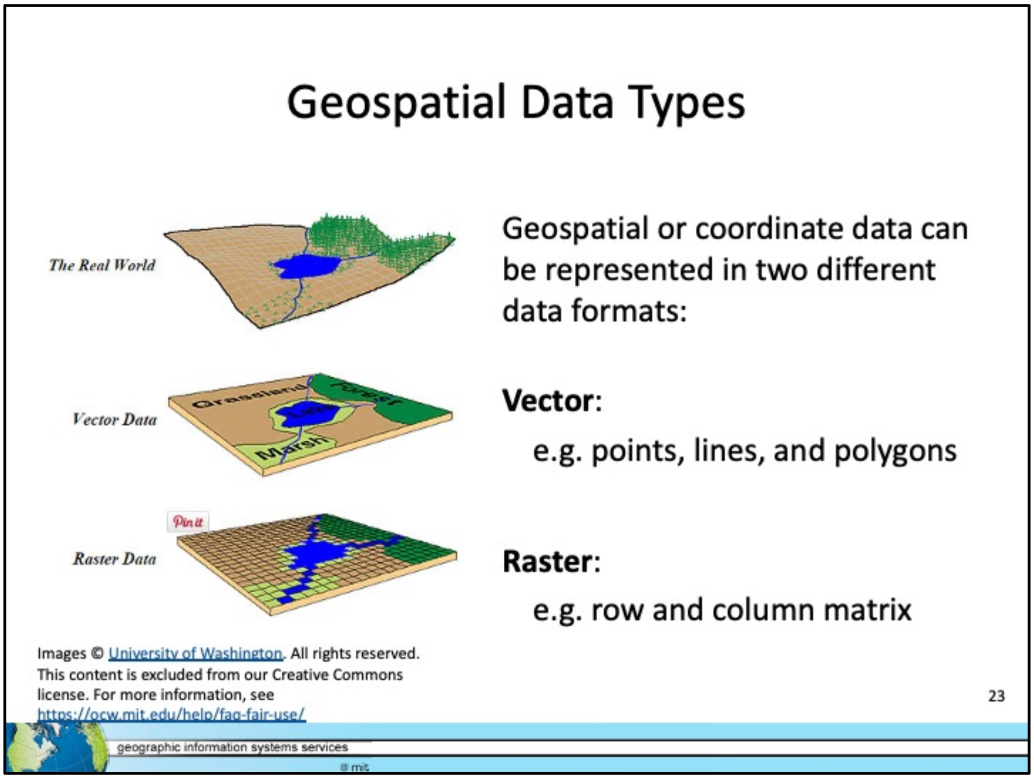


Answers: street networks, parks/open spaces, water, T stops, points of interest, etc.
Notes: data are organized by theme: infrastructure, hydrology, administrative boundaries, etc.

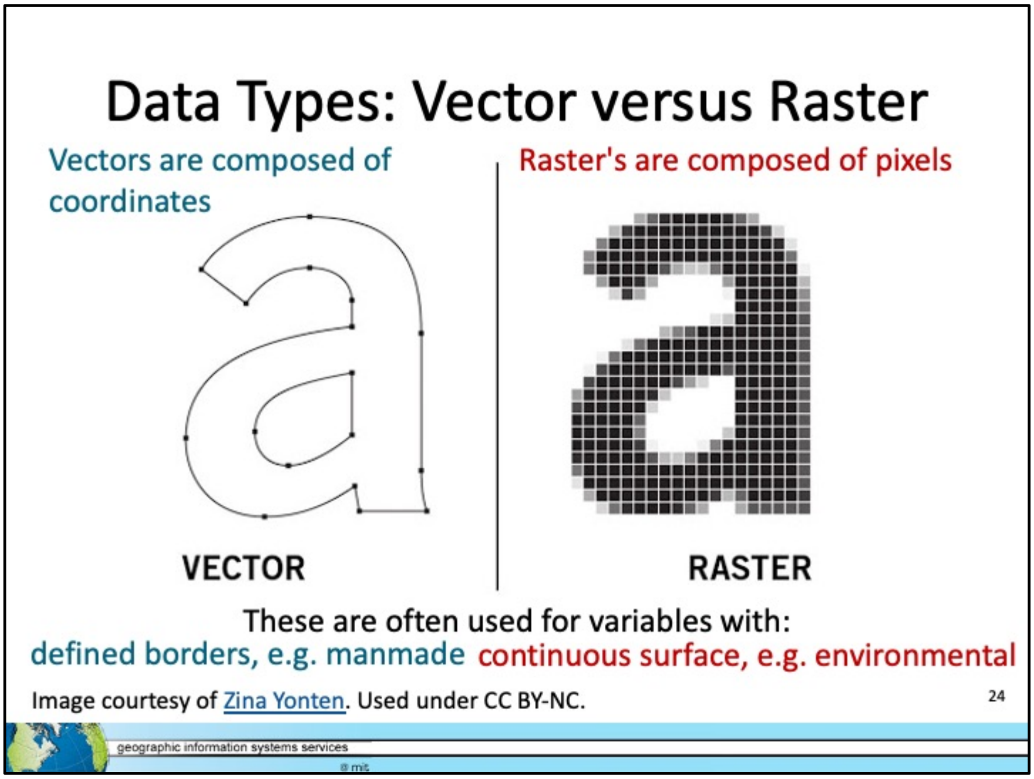
-When you create your own maps you can choose which data layers to include and how to visualize them to create a custom map.

MAPS & DATA:
SPATIAL DATA TYPES

22



There are two main approaches to representing real world data, vector and raster. Vector represents the world in points, lines, and polygons, while raster uses rows and columns.

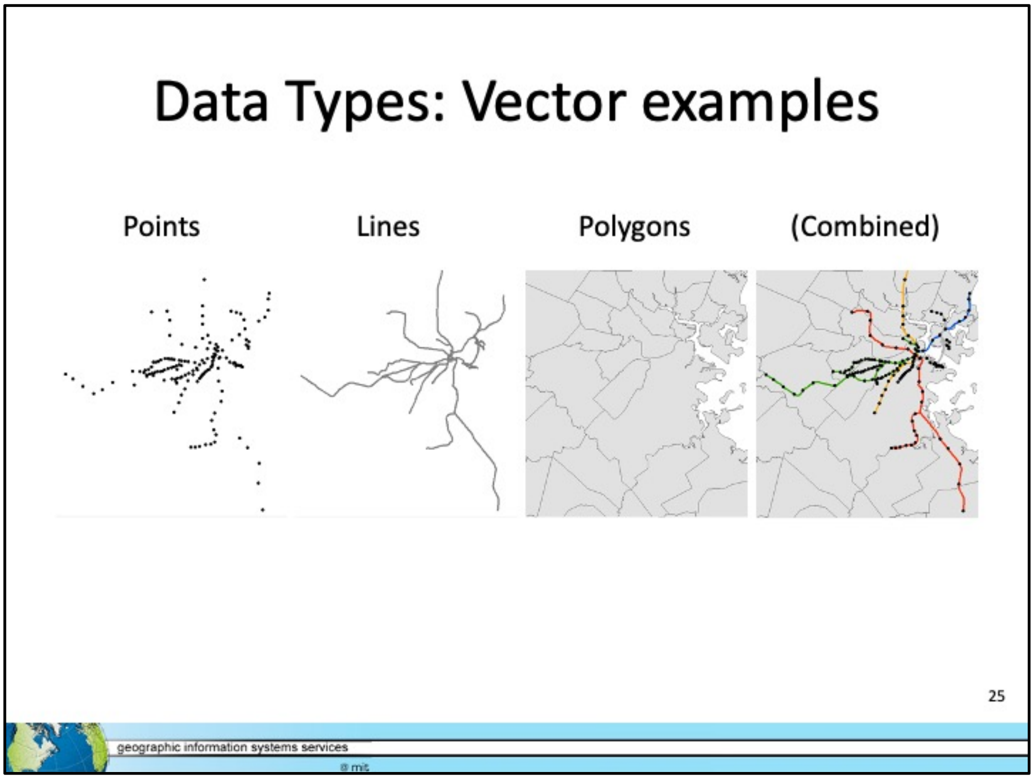


In some cases data can be represented as either a vector or raster, but usually certain types of data are better suited for a certain method of representation.

Vectors are composed of coordinates and are best suited to manmade features with defined locations and boundaries.

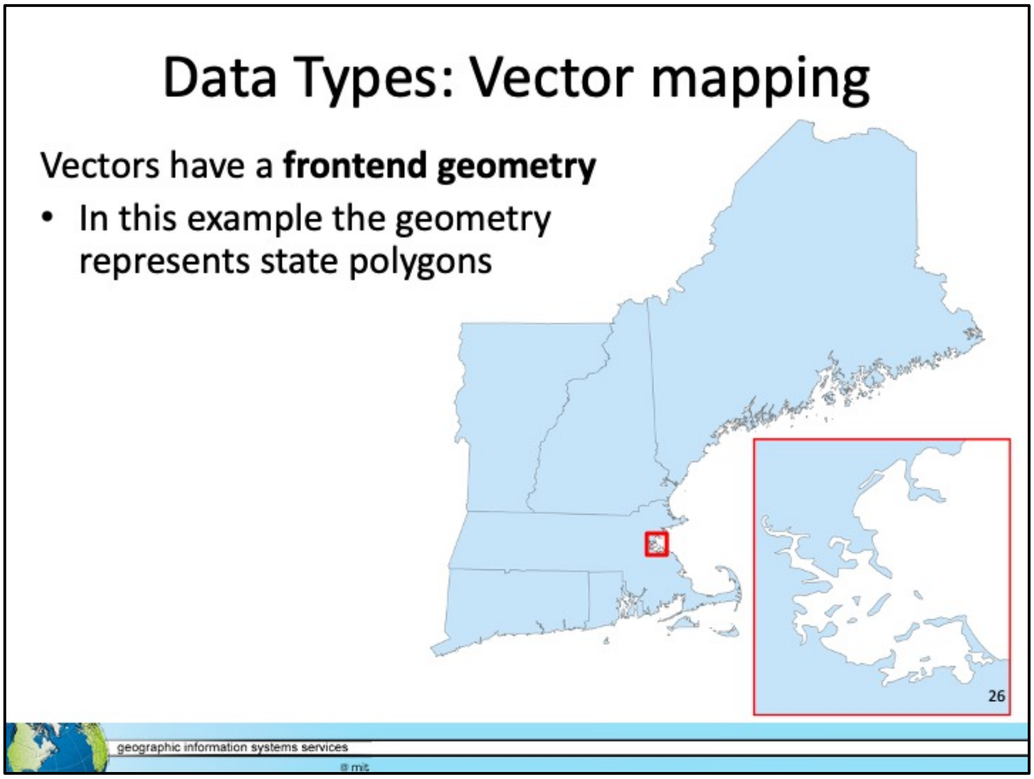
Rasters are composed of pixels and are best suited to variables, usually environmental, that change over surfaces such as temperature, precipitation, or elevation.

You are able to switch between vector and raster format to use different tools, which you can see in our GIS Level 2 workshop materials.



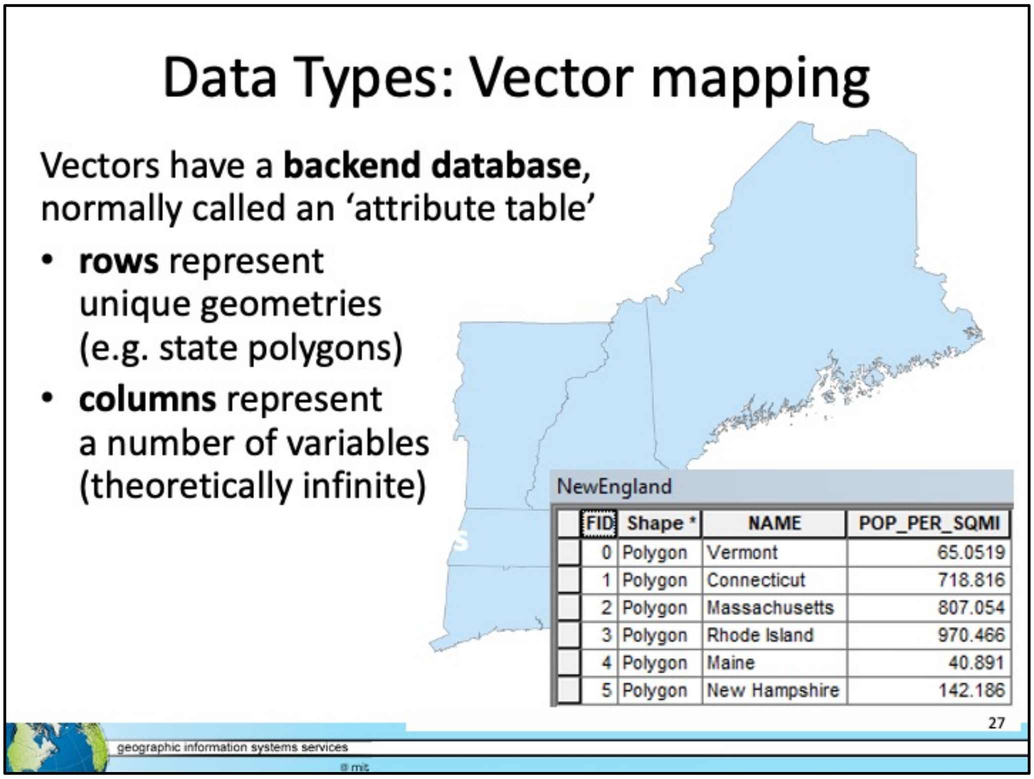
Here you see an example of points, lines, polygons, and all three combined

Question: Can anyone guess based on how these vector layers are color coded, also known as symbolized, what these datasets may represent?
Answer: MBTA Stops, Lines, and Towns, where the lines have been symbolized by color



Geographic data include both a frontend geometry, meaning what you see on the screen in GIS software and a backend database.

Vector data's frontend geometry is composed of coordinates and displayed as points, lines and polygons. Here is a layer with polygon geometry.

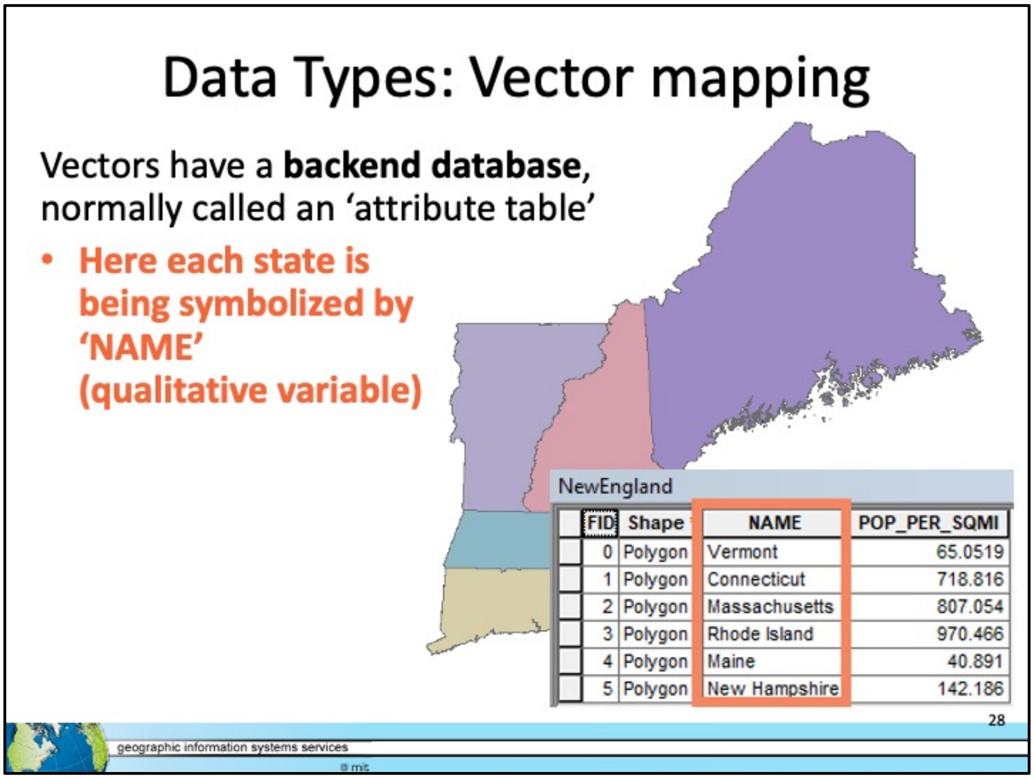


FID	Shape *	NAME	POP_PER_SQMI
0	Polygon	Vermont	65.0519
1	Polygon	Connecticut	718.816
2	Polygon	Massachusetts	807.054
3	Polygon	Rhode Island	970.466
4	Polygon	Maine	40.891
5	Polygon	New Hampshire	142.186

The backend database is called an attribute table. Each row is equivalent to one feature on the map. In this example each row represents a different state polygon.

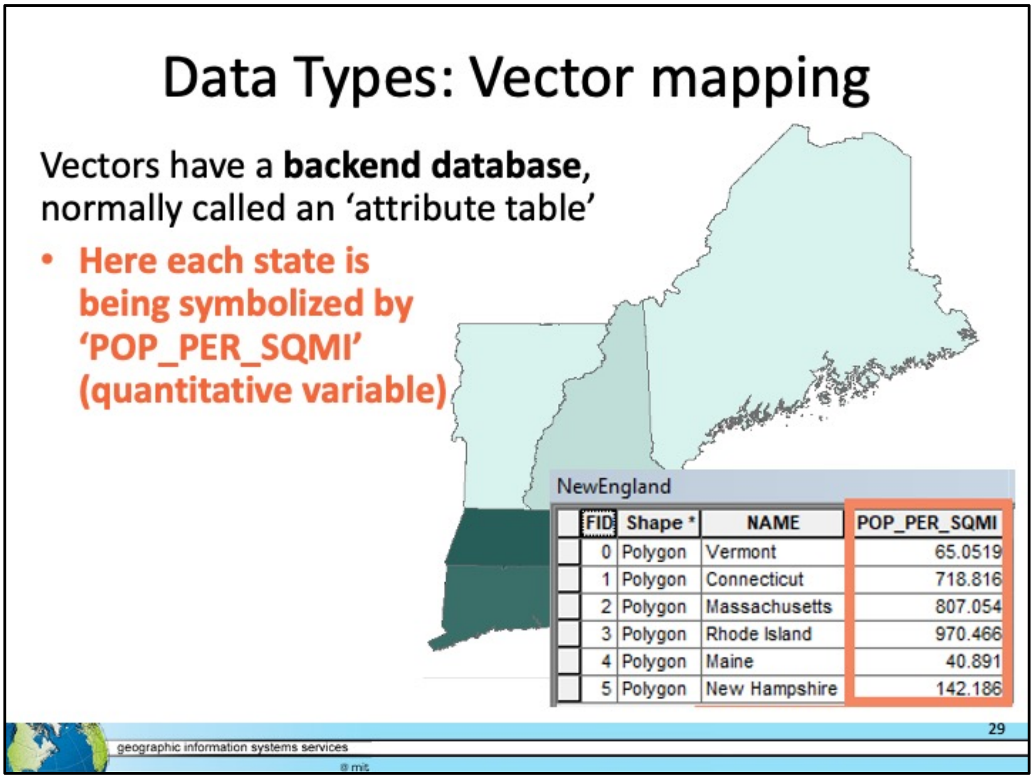
Each column is a different piece of information about that feature. In this example there is information about the state name and the population per square mile.

Vector data can have a large number of columns associated with their geometries, each of which can then be symbolized to produce different maps.

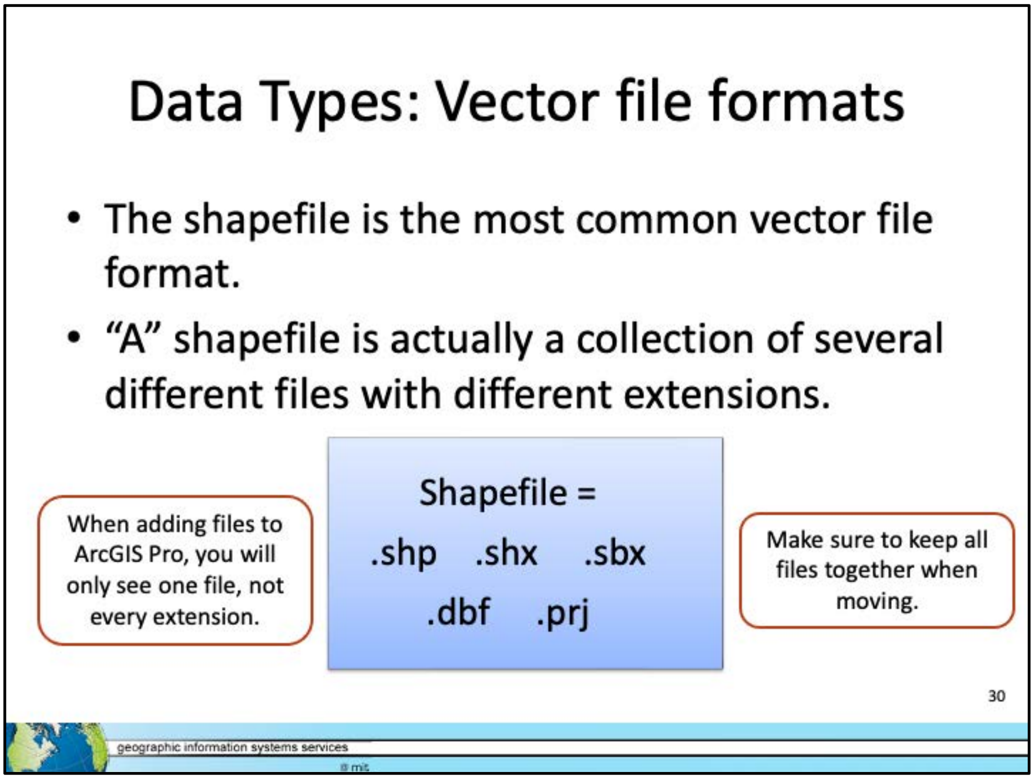


The map can be symbolized based on any column in the attribute table, meaning the color, size, shape, pattern, etc. of a feature can be changed to correspond to the data in a particular column.

Here the map was color coded based on a qualitative (aka categorical) variable, state name, where each unique state name was symbolized by a different color.



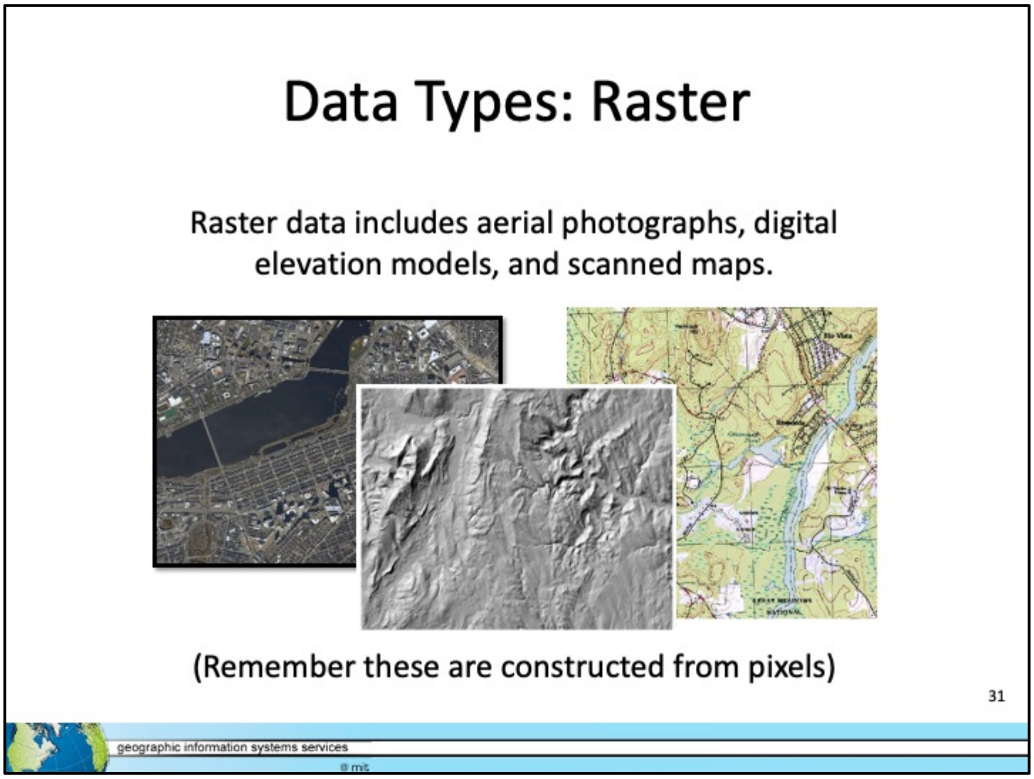
Here the map was color coded based on a quantitative (aka numerical) variable, population per square mile, where each class of values was symbolized by a different color. We will talk about classes in depth later in the presentation.



A shapefile is a open source format for vector data that can be opened in any GIS software.

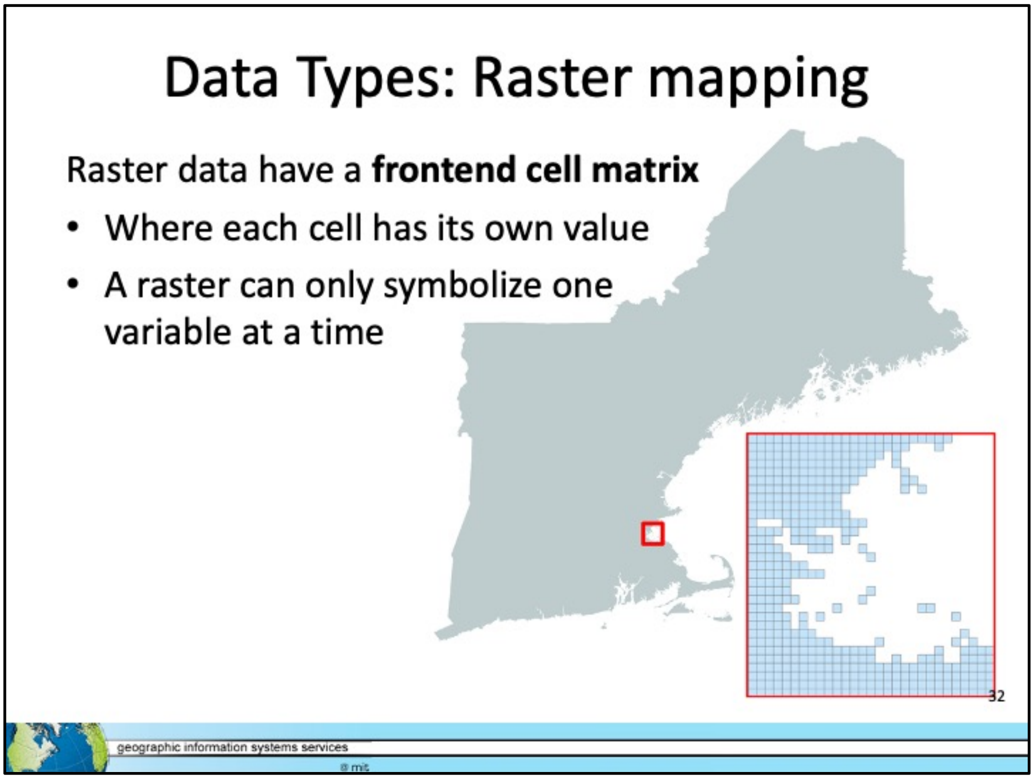
Shapefiles are often in a zipped (.zip) folder because they include several different files. This folder needs to be unzipped to use in ArcGIS Pro, but files can be imported directly from the .zip folder in QGIS.

The .shp includes the actual geometry of the data, the .dbf includes the attribute table, and the .prj contains the map projection, which is covered more in GIS Level 2. Other files include indexes that speed up the loading and display of the data. Keep these files together when you move or share data in order for them to load properly.



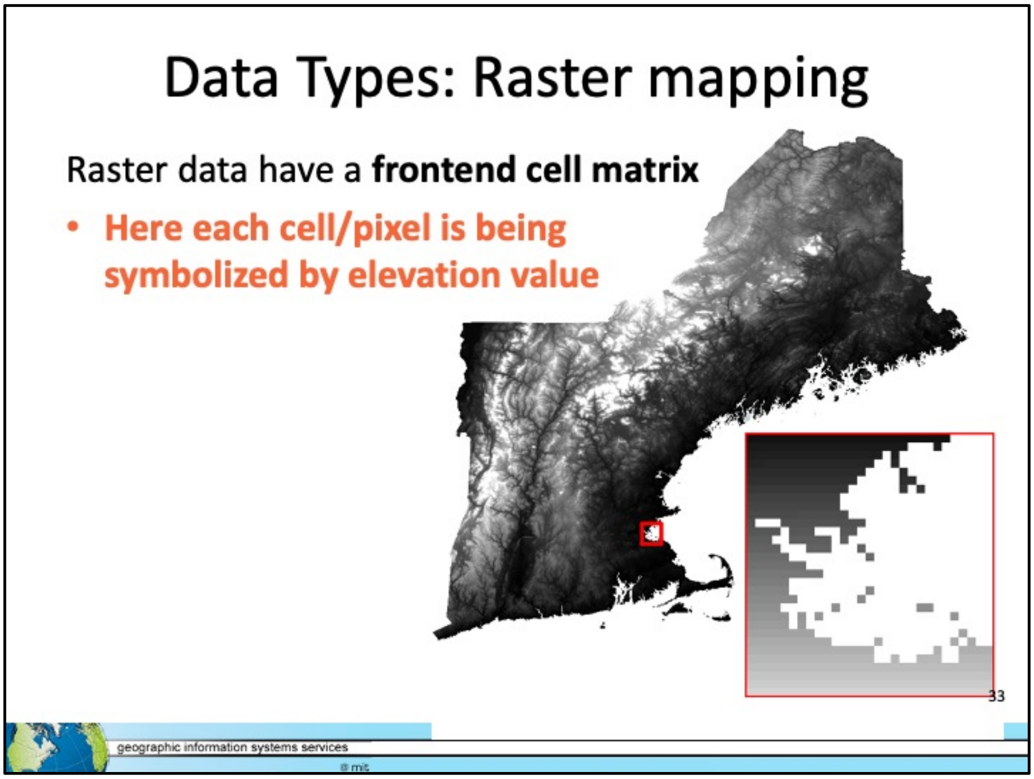
Here you see examples of rasters, such as aeriel photographs, digital elevation models, and scanned maps. All of which are constructed from pixels.

Additional Notes:
Early maps were created from surveys and early digital geospatial data were “digitized” from these maps. Data are now created using GPS. Some data are created from aerial photographs. Data are constantly being updated

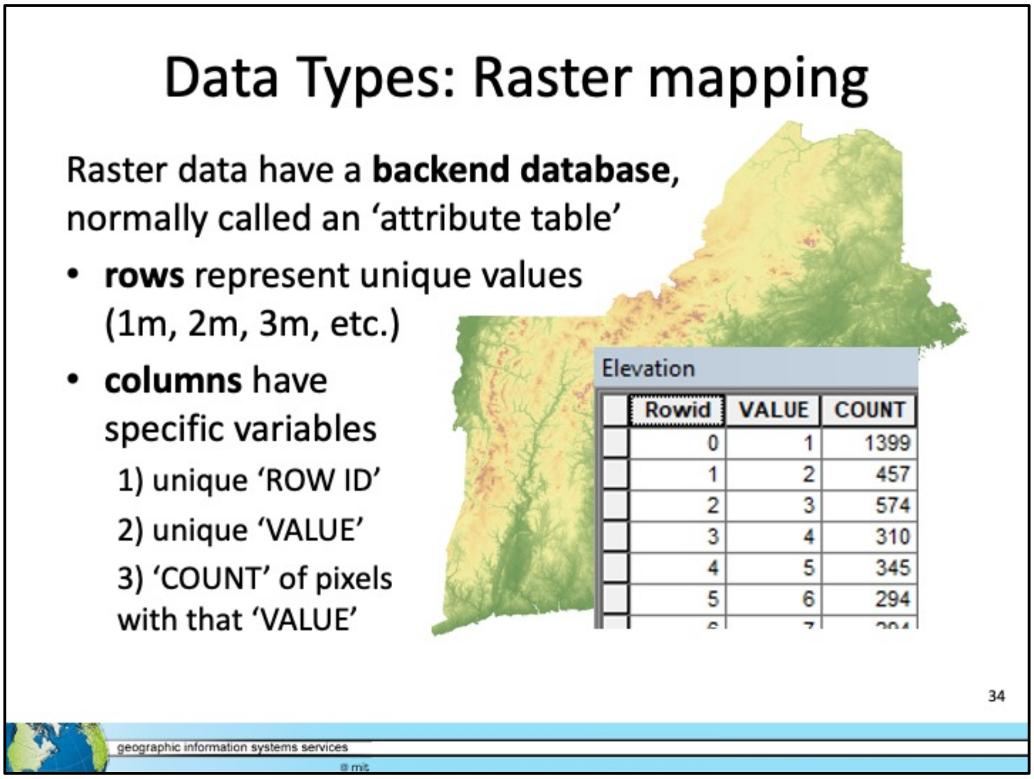


Raster data is a continuous cell matrix. Each cell or pixel is the same size and has its own value.

Rasters can only symbolize one variable at a time due to how its attribute table functions.

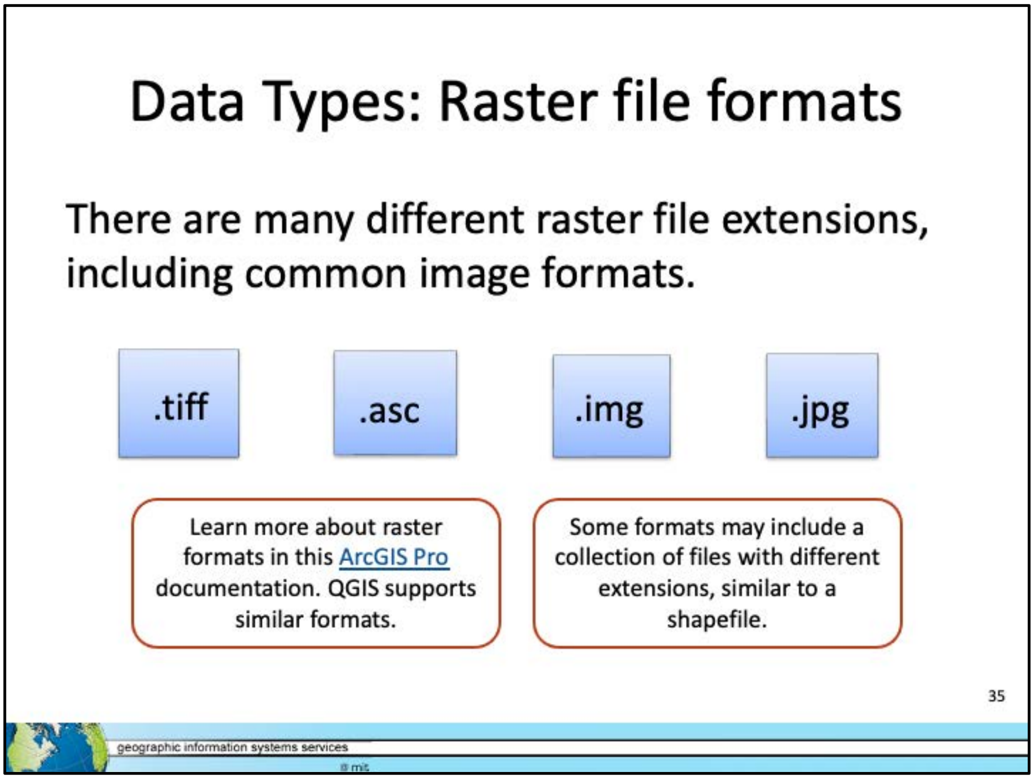


Here the map is color coded based on a quantitative (aka numerical) variable, elevation, where each unique pixel value is symbolized by a shade of grey stretching from black to white.



Raster data have attribute tables with specific properties: a unique id for each cell/pixel, the value of that cell, and the count of other cells with that same value.

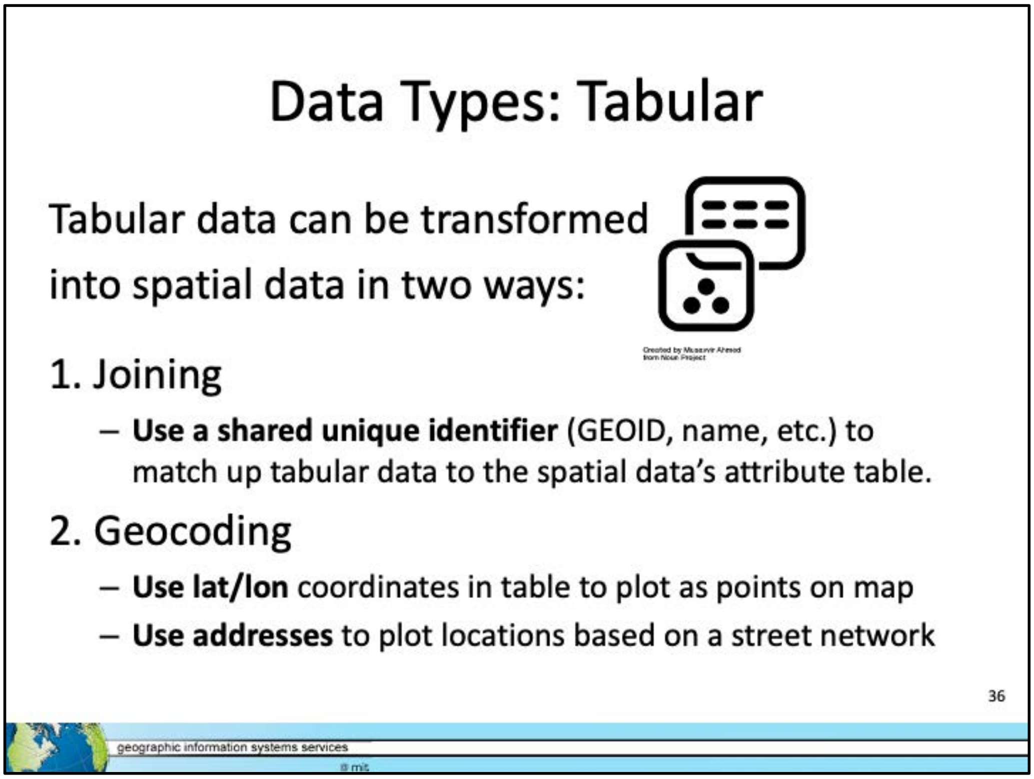
Note: the map is the same elevation data symbolized with a different color ramp.



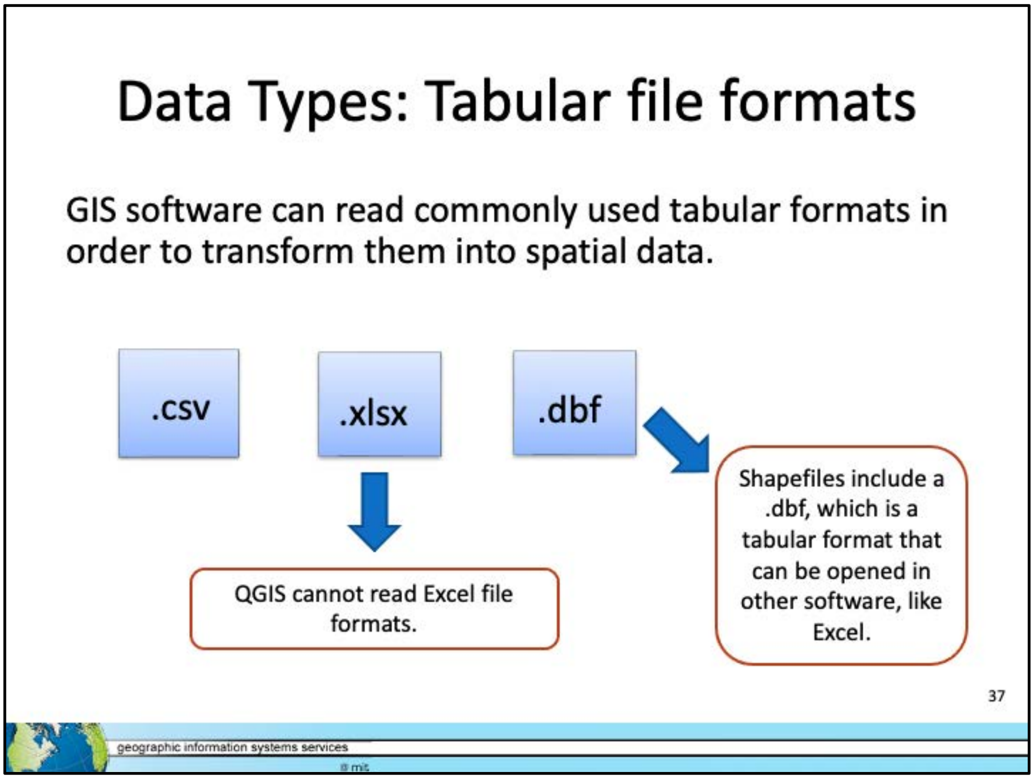
Raster file formats include common image formats and there are many more raster file formats that are not listed here.

There are often associated files that tell the GIS software where to place the raster on the map, similar to how a .prj file works in a shapefile.

If you import an individual image, such as a .jpg of a scanned map, you will need to do what is called “georeferencing” and tell the software how to align it with the rest of your data.



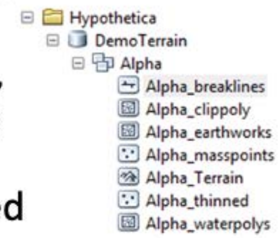
You can convert tabular data such as those in spreadsheets so long at the data contains certain geospatial information (i.e. shared unique identifiers, lat/lon, and/or addresses). See our GIS Level 2 workshop materials for hands on experience.



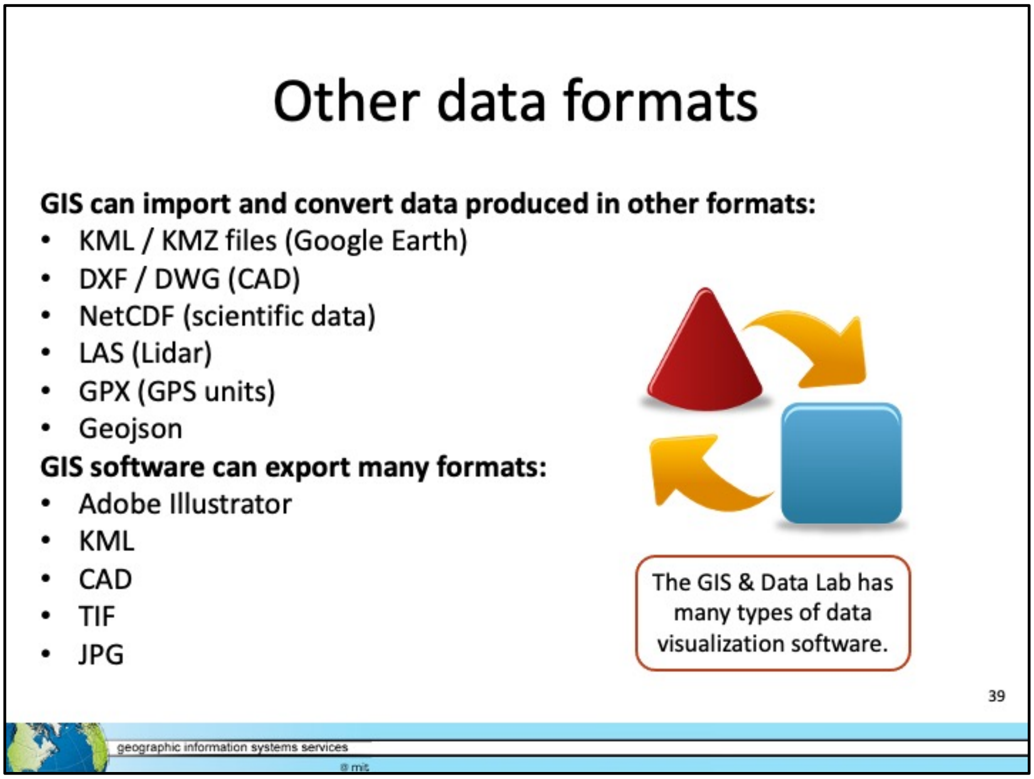
GIS software can read common tabular data formats. If there is geographic information included in the data table, GIS tools can be used to transform the table into a shapefile.

Geodatabases

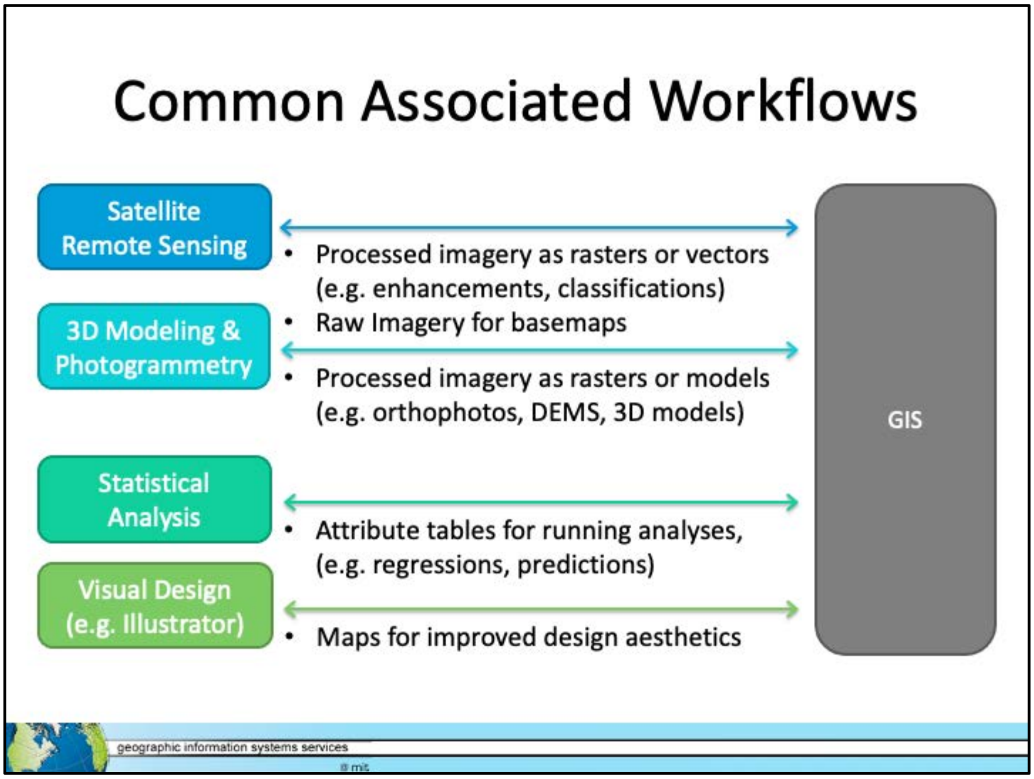
- ESRI/ArcGIS storage system
- a collection of geographic datasets of various types held in a common file system folder
- **Advantages:** larger files size limits, faster processing time when using analysis tools
- **Disadvantages:** can only be opened in ESRI software
- Learn more about using geodatabases in Pro.



Geodatabases are a file storage format used in ArcMap and ArcGIS Pro.
Geodatabases are similar to zipped files in that they can store and compress a variety of different data types, including vectors, rasters, and data tables.
They are useful for organizing data and speeding up processing time when working with large files. A disadvantage is that they can only be opened in ESRI software.



GIS software can import and export data in a variety of formats. Some common import formats include KML/KMZ from Google Earth and CAD files. Maps can be exported in image formats for reports or presentations, such as CAD or Illustrator for further development.



Many people do not just need GIS software to conduct their research.
The ability to import and export data and maps allows you to work in a variety of software before or after using GIS.
Common workflows include using GIS in conjunction with remote sensing software, which is used to analyze satellite imagery, with 3D modeling software such as CAD or Rhino, using attribute tables in statistical analysis software, and creating elaborate maps designs with visual design software.

Exercise 1

- Goals:
  - Become familiar with the GIS interface
  - Learn how to add data
  - Explore data types & attributes
- Complete either the QGIS or ArcGIS Pro exercise from your workshop folder.

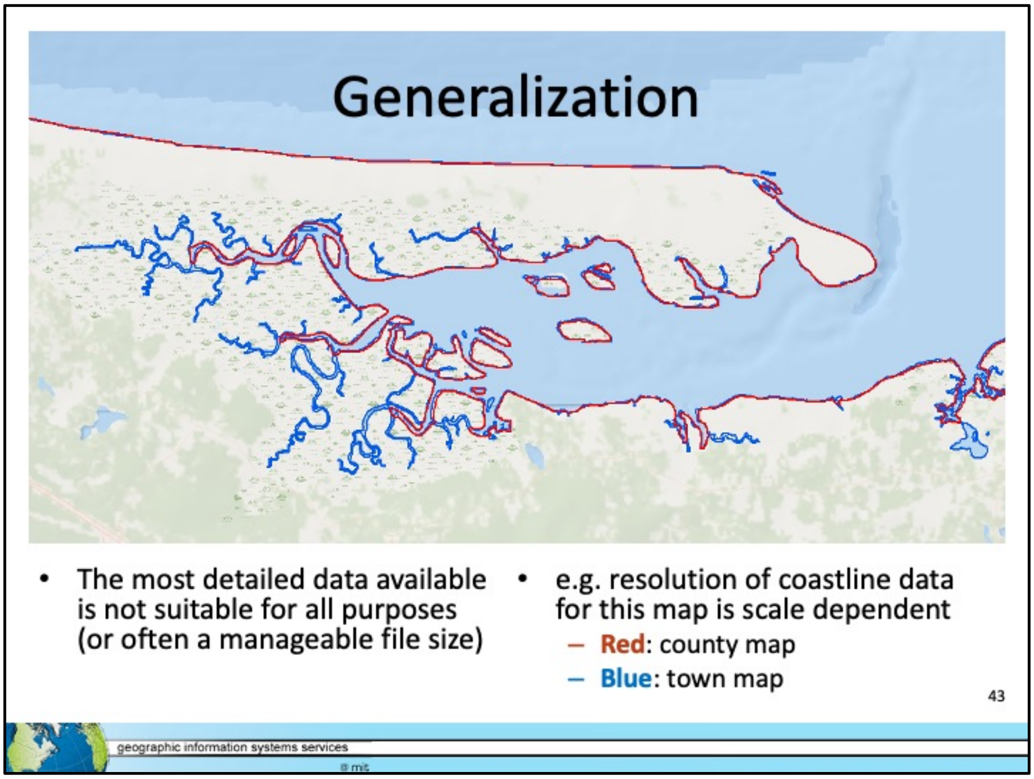
41

geographic information systems services

@ mit

MAPS & DATA:
CHARACTERISTICS OF SPATIAL DATA

42

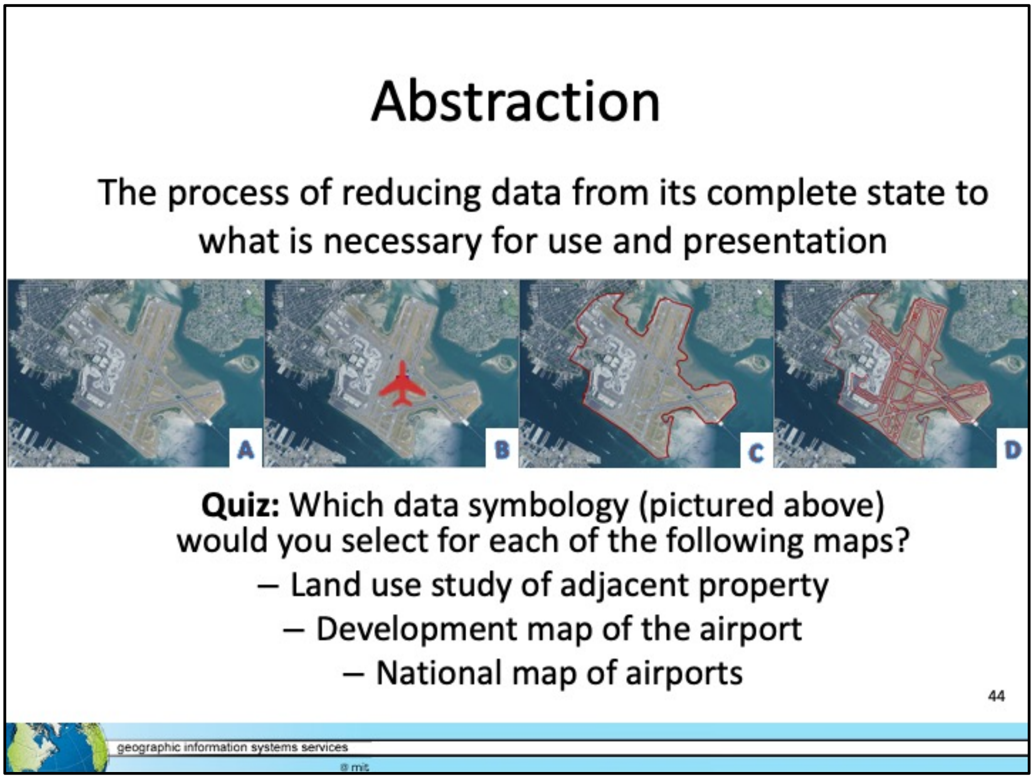


There are characteristics of spatial data that make it unique from other types of data. You need to know about these special features in order to find and use spatial data.

Spatial data is generalized, meaning it is simplified from what you would find in real life. The more detailed your data are, the larger the files sizes, which means more data to store and longer processing times when analyzing it in GIS software.

Depending on your project, you may need data with more or less detail. In this example, the coastline outlined in red would be suitable for display on a county or state map. It would be difficult to see a lot of the small inlets and islands at that scale so you want something more generalized, with less detail.

The coastline outlined in blue would be suitable for a map of the town or something smaller, such as a specific bay or beach. More detail is often useful when mapping or analyzing a small area.



Spatial data are also abstracted, meaning they include only what is necessary for your map and analysis. It would be impossible to include every feature that you see in real life on a map. Not only would it create large files, but the map would be difficult to read.

This example includes data that have been abstracted in different ways, for different purposes.
Example A is satellite imagery of an airport without any additional symbology.
Example B uses a symbol of an airplane to represent the airport.
Example C uses a polygon to represent the border of the airport property.
Example D uses polygons and lines to represent the airport border and the runways.

Which data symbology would you select if you wanted to do a land use study of properties adjacent to the airport?
C - Because it shows the border, you can easily see what is adjacent. Example D also shows the airport border, but includes runway information, which you do not need.
Which data symbology would you select if you wanted to create a map of potential new develop within the airport? (pause)
D - Because it shows the airport layout. It would be important to know where the runways are when planning future development.
Which data symbology would you select if you wanted to create a map of all airports in a country? (pause)
B - Because it just shows the airport as one symbol. If you used the symbols pictured in C or D, you map would be messy and difficult to read.

Spatial Resolution/Scale

Suitable data geometry is dependent on scale:

e.g. roads are polygons at local scale but lines at national scale

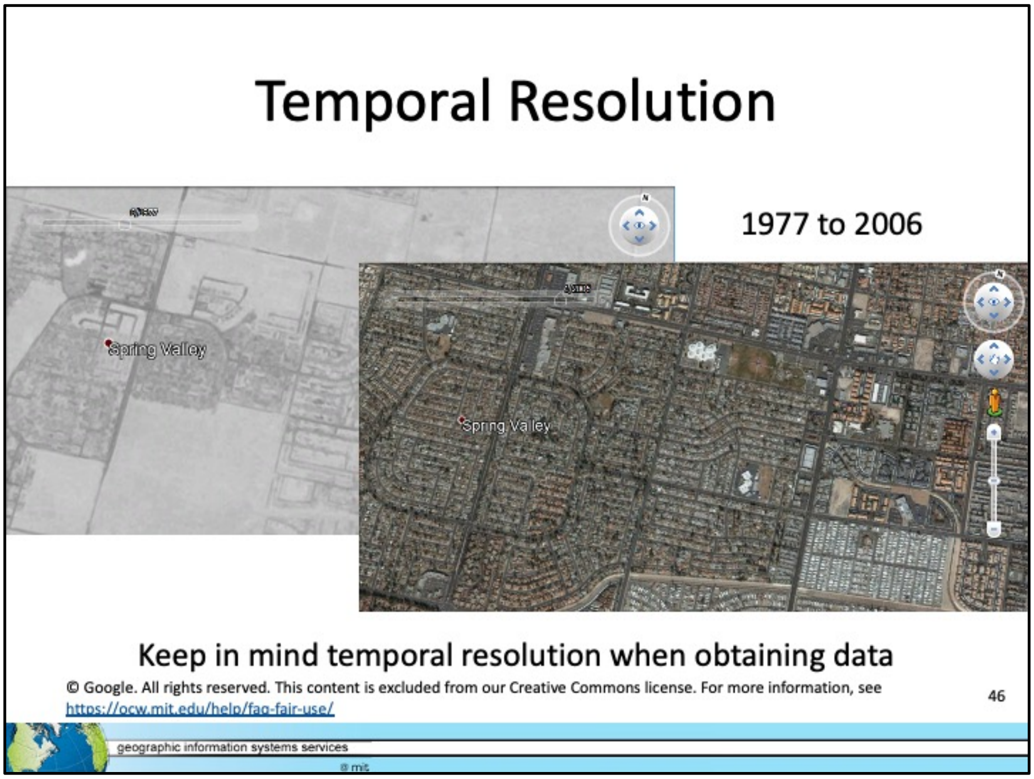
45

geographic information systems services

The spatial data you make visible on a map depends on its scale, meaning how small or large of an area you are showing. You should show enough detail at a particular scale so that your viewer can clearly see all the features you have added.

Question: Can you spot any other features dependent on scale?

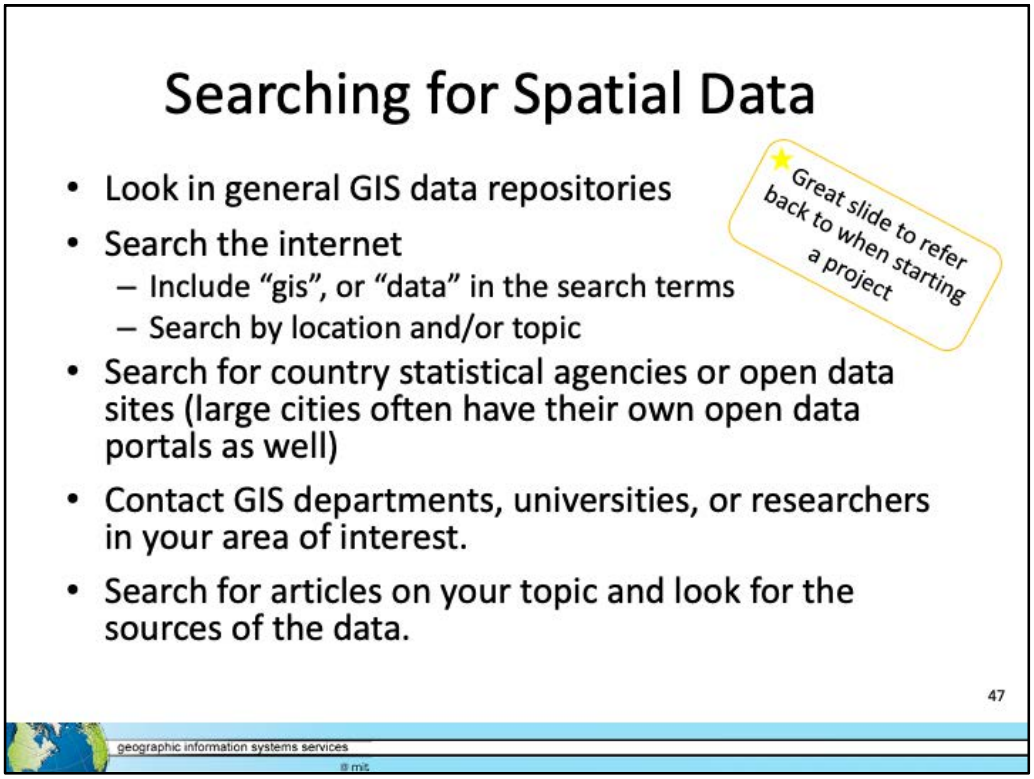
In the city map on the left, you can see points of interest, street widths, directions and names, and public transportation stops. These are not visible in the regional map on the right because they are not needed at that scale and would make the map impossible to read. The regional map includes points that represent major cities, highways, and larger state and national parks.



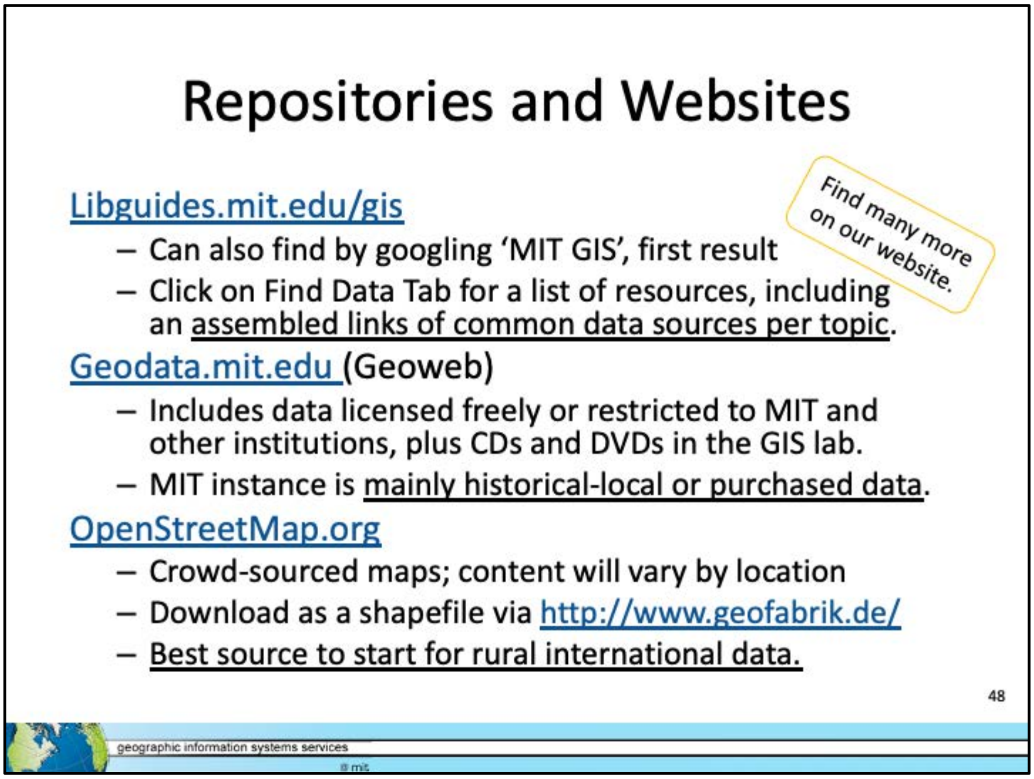
Spatial data changes over time. The data pictured on the previous slides are only accurate for that particular point in time. Coastlines may erode or be created in a storm. An airport could expand or close. The names of stores or number of streets may change.

In this example, you can see from the imagery that Spring Valley had a lot of development from 1977 to 2006. Data for roads or houses will look different at different points in time.

Note: These aerial images are from the Google Earth Historical Imagery Time Slider.



Now that you know what type of spatial data to look for, where can you find it?
A lot of data is available freely online, especially for the US.
These are tips you can use in the future for data searches.

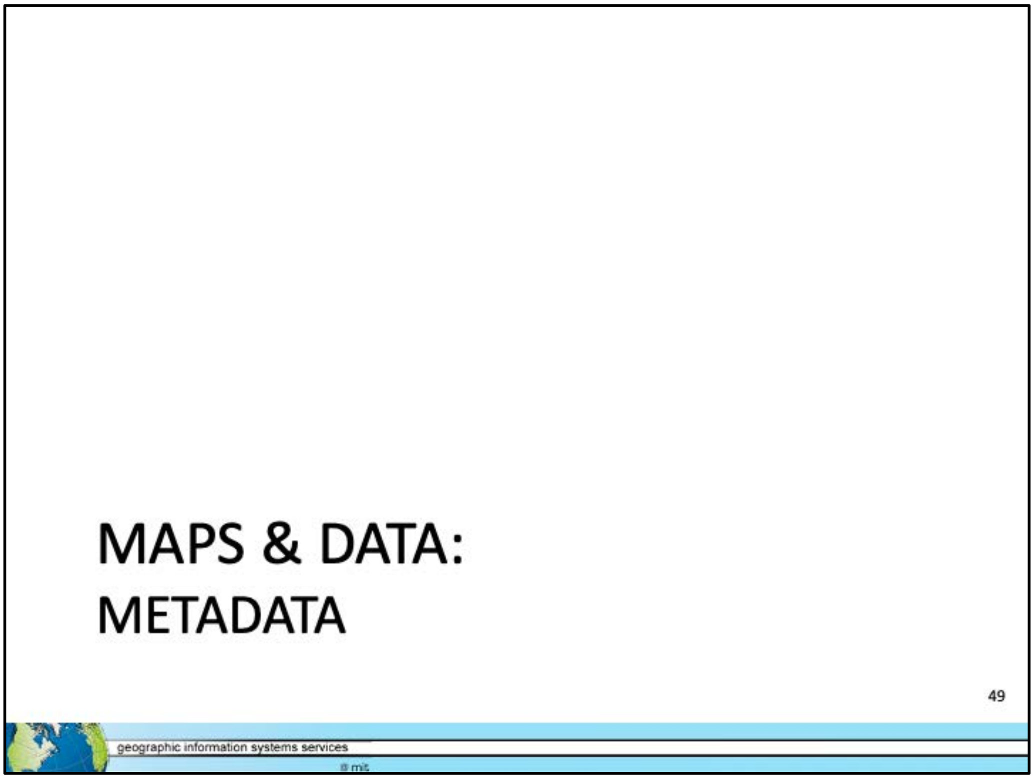


Here we've listed 3 of the data repositories we frequently use when assisting researchers.

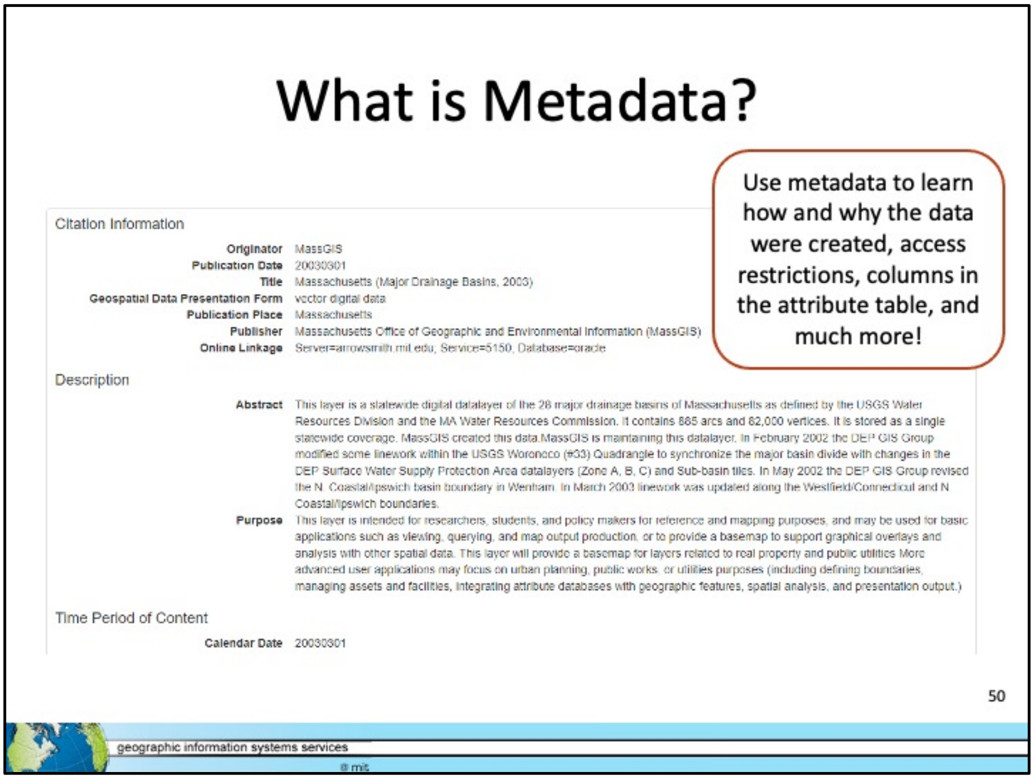
At the top is the MIT Libraries GIS research guide which you can find by simply Googling MIT GIS. If you hover over the Find Data tab, you'll see a breakdown of GIS data sources by geography and subject.

MIT also has a resource called Geoweb which includes freely available data as well as licensed data restricted to MIT and other universities. The data restricted to use by MIT community members consists largely of data purchased by the GIS team on specific topics or areas of the world.

OpenStreet Map is an open repository of crowd sourced maps. The amount of GIS data will vary by location.



Now we will talk about metadata, aka information about our data.



Metadata is a way of describing an information resource so you can better understand the data. It often describes how and why a dataset was created as well as provides information about any codes used within the dataset.

Metadata Examples

1. MassGIS:
   https://www.mass.gov/info-details/massgis-data-marine-beaches
2. GeoWeb: geodata.mit.edu/catalog/mit-w37ehgh6nvl4w
3. City of Boston: https://data.boston.gov/dataset/traffic-signals

51

geographic information systems services

@ mit

Shown here are some common metadata sources for geospatial repositories around Boston. Let's go through each link and see how their metadata are represented differently.

MAKING GREAT MAPS:
DATA VISUALIZATION PRINCIPLES

52

geographic information systems services
@ mit

Let’s talk about some data visualization principles for making great maps.

Making Great Maps

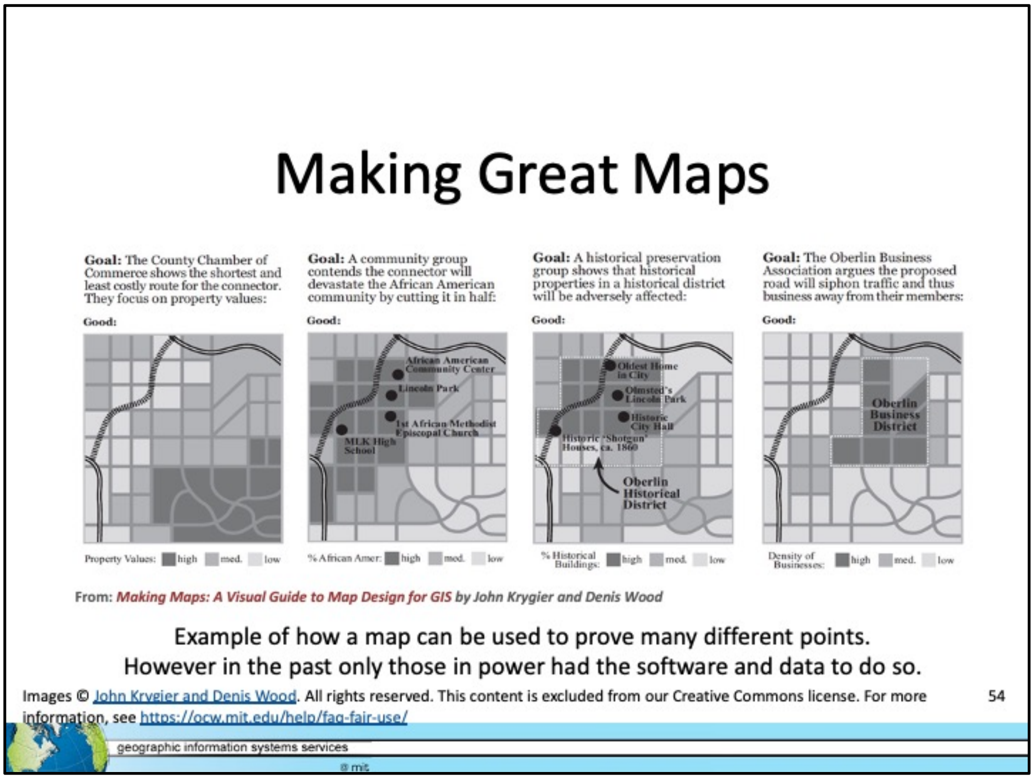
- Cartography is the art and science of making maps
- Maps are always simplifications of reality, which makes them helpful when making decisions or explaining patterns
- Maps are designed by people (who have intentions), so we have to create them responsibly

53

geographic information systems services

@ mit

Art & science = design & analysis
Simplifications of reality = you can't show everything
Designed by people = there's a motive behind their creation



In this example, we see successive maps all using similar symbology but different variables to argue the location a highway connector should be placed. Things to keep in mind are:

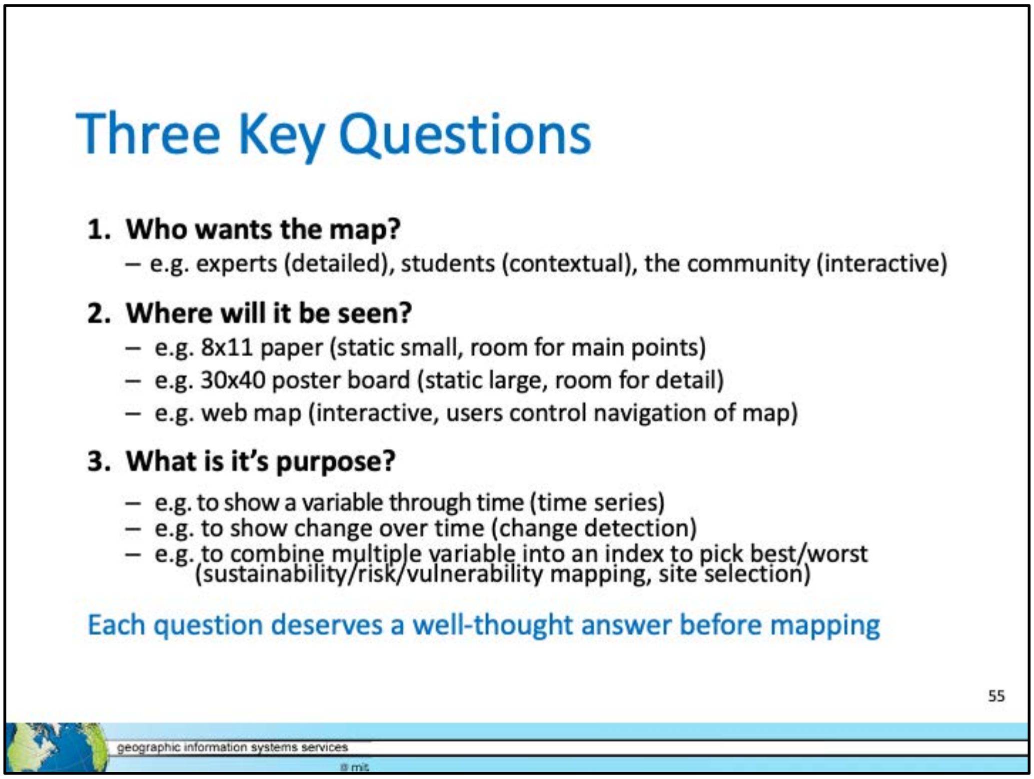
-only relevant features (for a particular group) were selected to be in each map.

-each map is by a different creator and trying to convey a different message

-each map is good for a different group

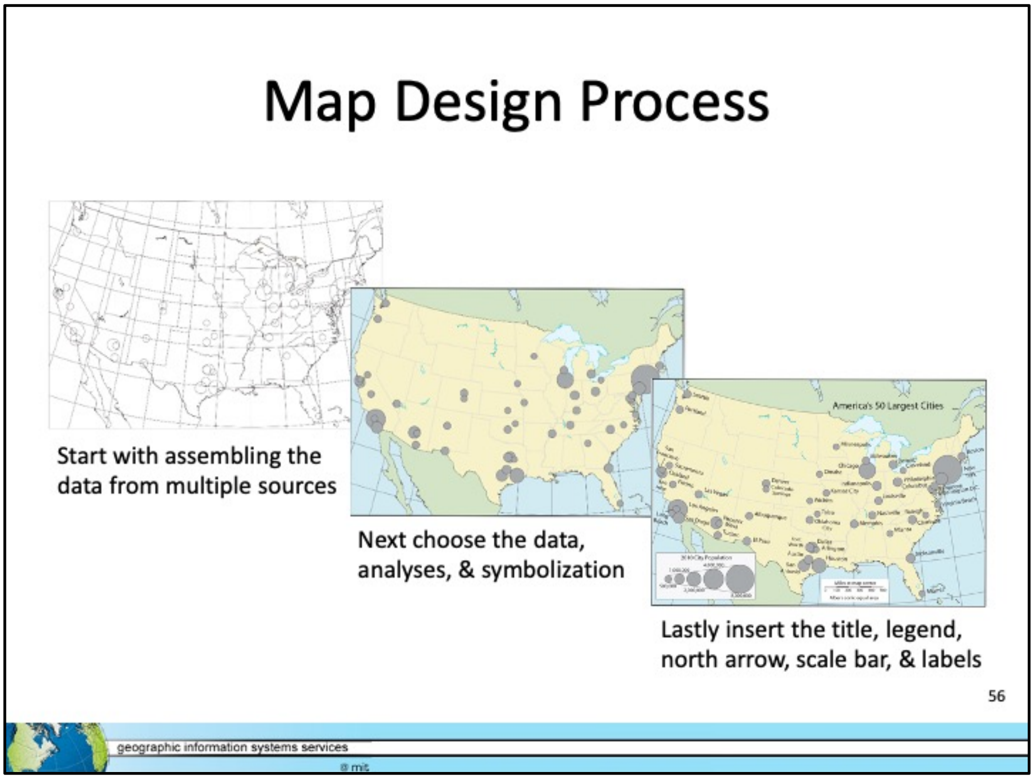
Question: which is best, and what they think the solution should be?

Answer: (all correct, but all biased)

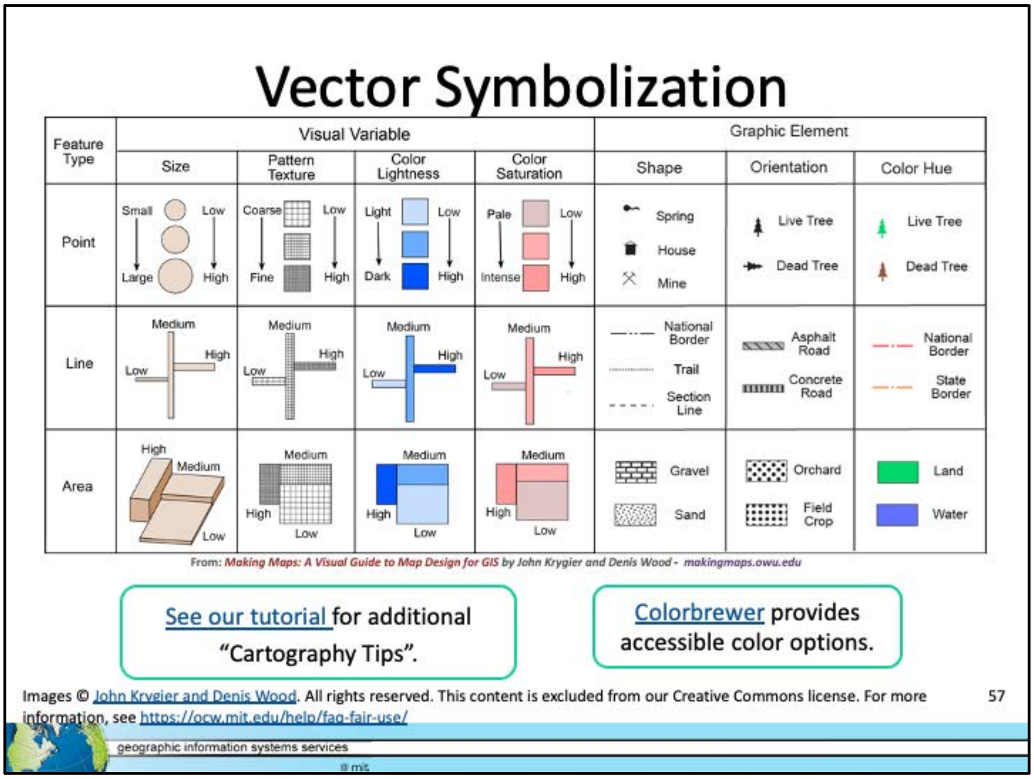


You also want to keep in mind some key questions, such as who wants the map and where it will be seen, both of which will affect the level of detail and whether or not the map should be interactive.

Depending on the purpose for the map will affect what type you will create, such as change through time vs. space, and/or combining multiple variables for decision making.

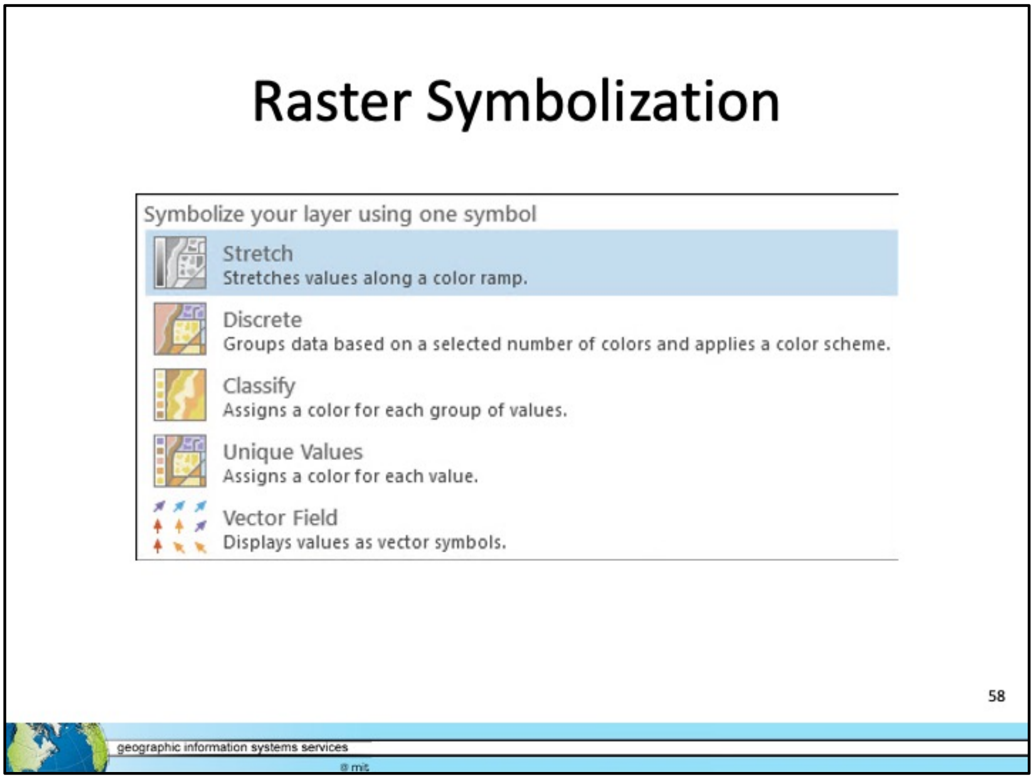


Stage 1: collecting your data (we talked about this at the beginning of the workshop)
Stage 2: symbolize your data (we will talk about the next 2 items now)
Stage 3: create a layout (add title, scale bar, legend, north arrow, etc.)

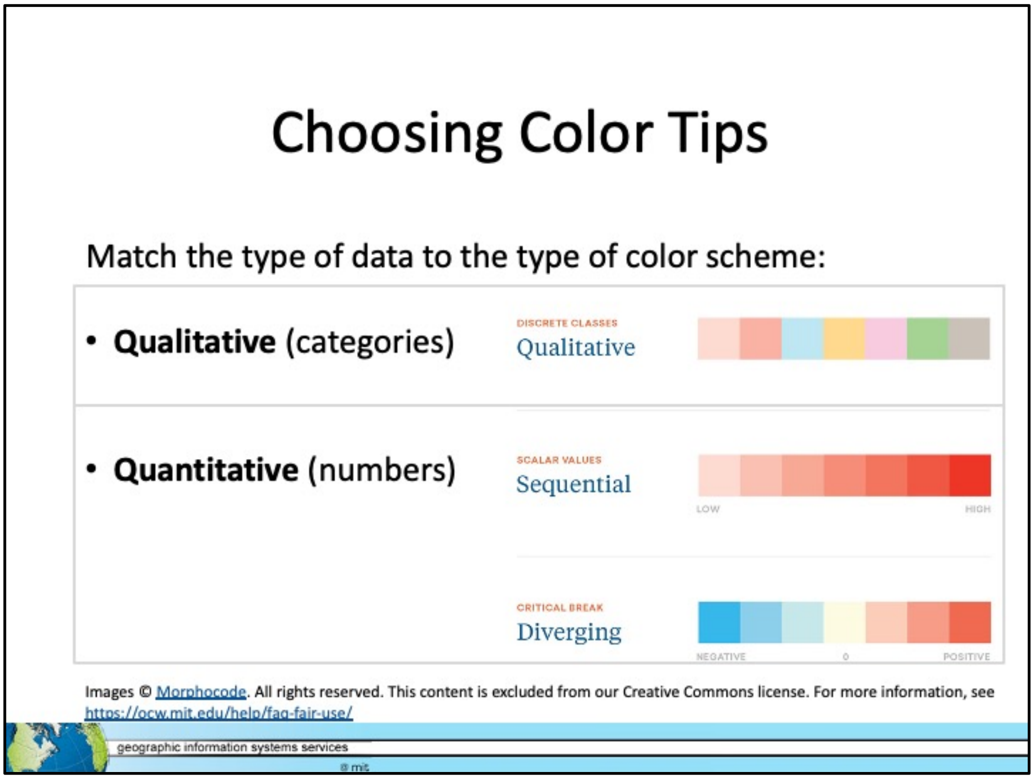


There are many ways to symbolize vector data as points, lines, and polygons. While color is one of the more common ways to symbolize data, this chart shows additional ways to show differences among features on your map by varying size, pattern, shape and orientation.

When choosing any type of symbology, it is important to think about accessibility, especially in relation to color blindness. The Colorbrewer website allows you to filter colors based on those that can be viewed by people with color-blindness. ArcGIS Pro also has this option in its symbology menu.

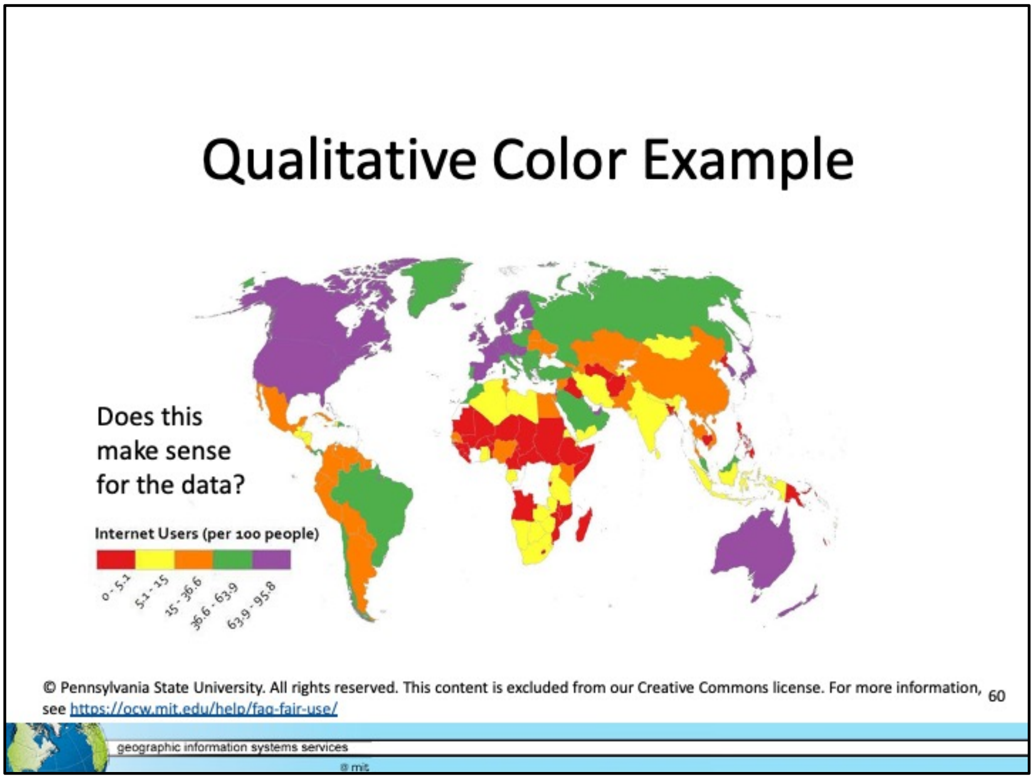


Raster data are symbolized differently from vector data, as you saw in the earlier modules about types of spatial data, because they are continuous surfaces. Raster data can be displayed by showing all data values, grouping values into categories, varying colors across the surface based on the value, or creating a vector field using symbols, which you might see in a map of wind direction and speed.



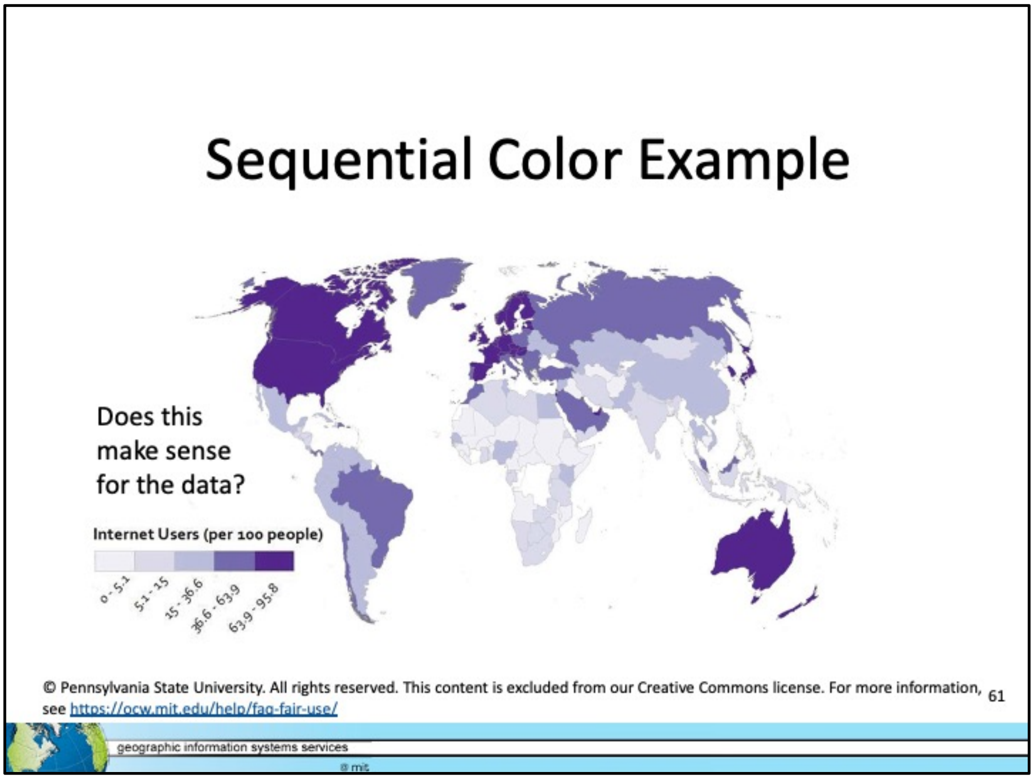
When it comes to choosing color, you want to be mindful of they type of data you are working with. Qualitative data often uses different colors for each category, while quantitative data often uses one color or two if there is a diverging phenomena.

Examples of each would be:
Example of qualitative/categorical: land use
Example of sequential numbers: population
Example of diverging numbers: weather/political (e.g. red, blue, neutral)

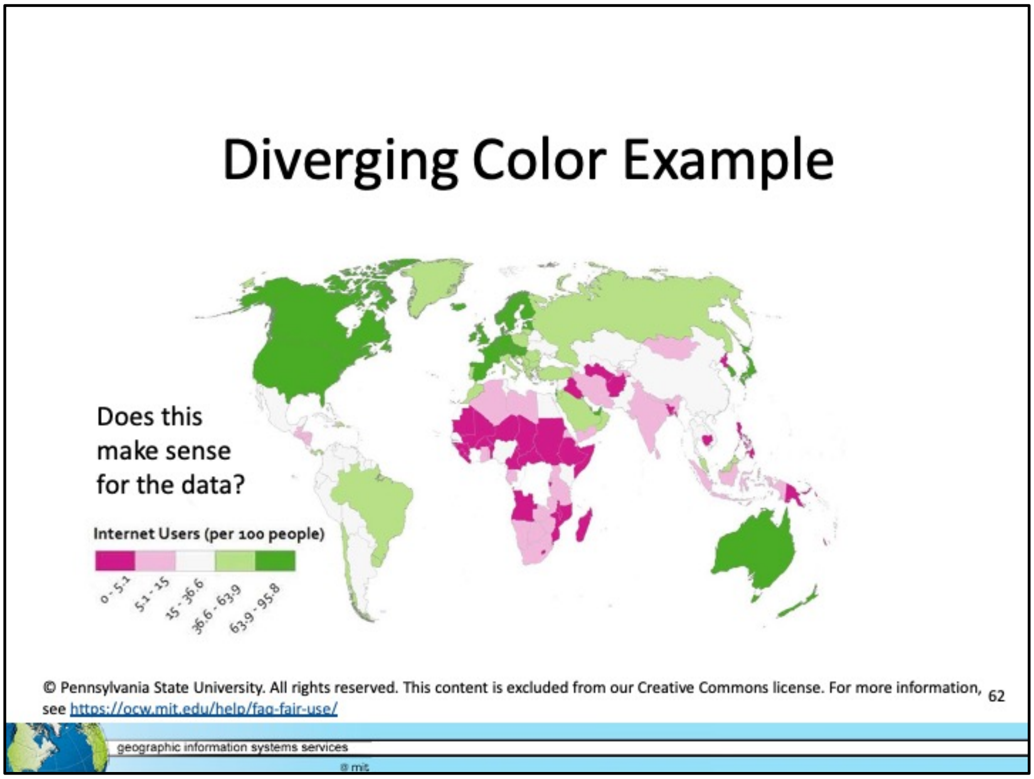


Let’s test our knowledge. In this example we have the variable “Internet Users (per 100 people)”.

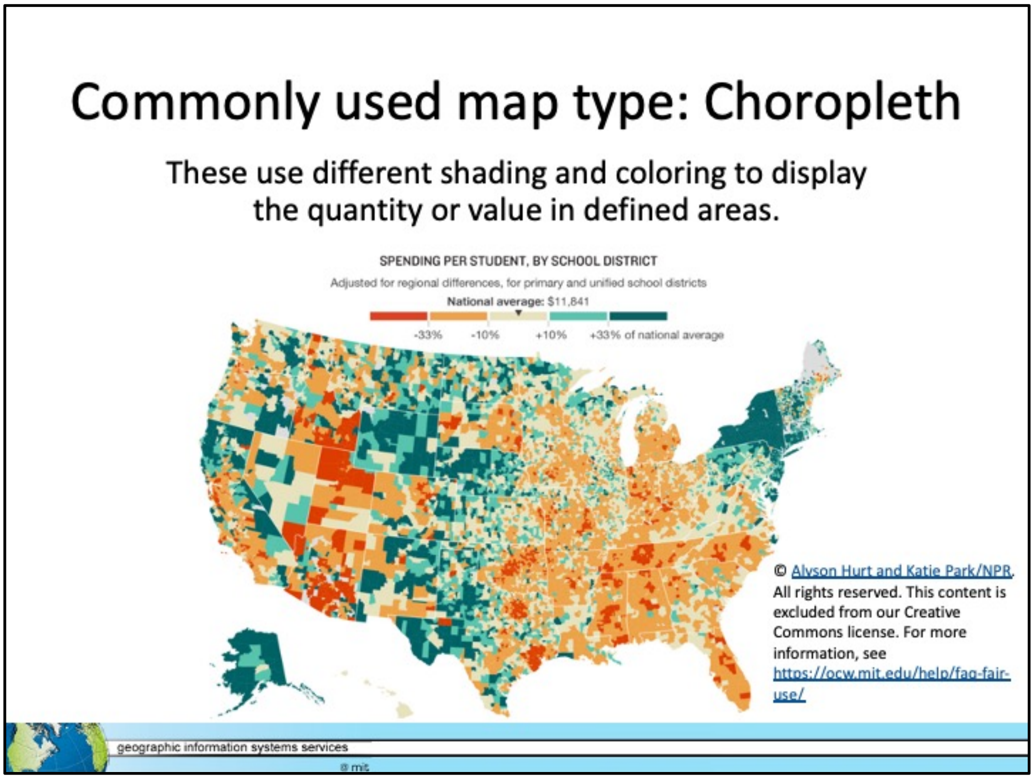
Question: Can you tell where the highest internet users are?
Answer: No, doesn’t make sense with quantitative (numerical) data.



Question: Can you tell where the highest internet users are?
Answer: Yes, this has an ascending trend which is reflected in the darker color indicating greater intensity.

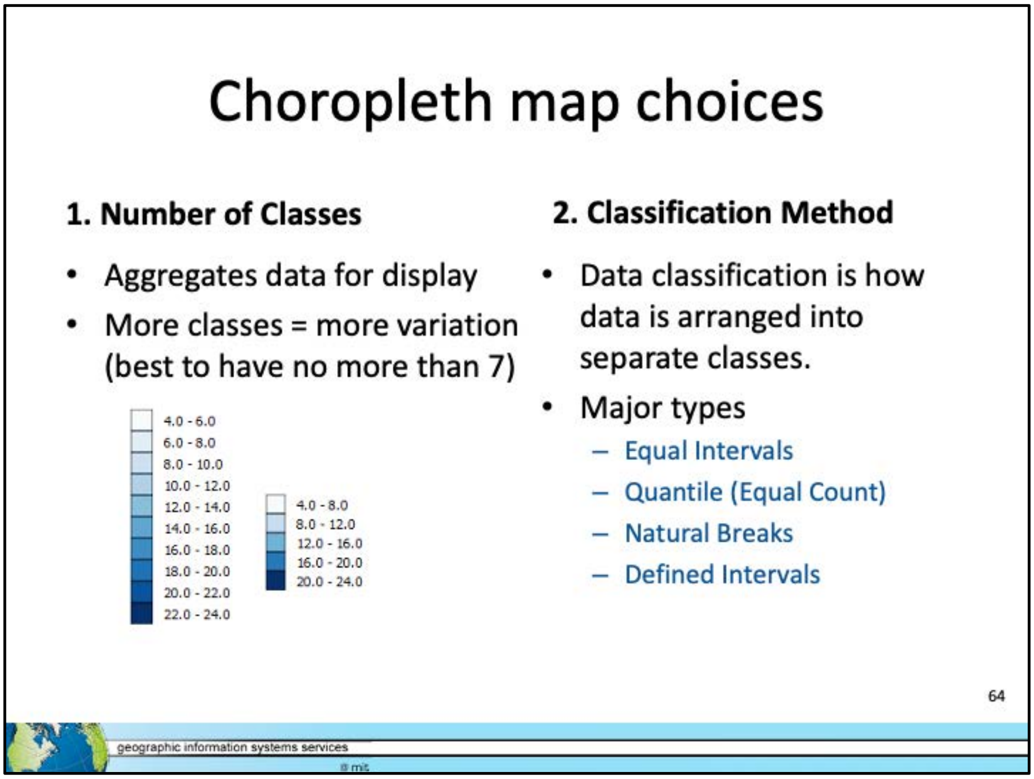


Question: Can you tell where the highest internet users are?
Answer: It depends. Potentially, this would make sense if there was a certain phenomena that occurred and you wanted to show above/below this value (white colored area).



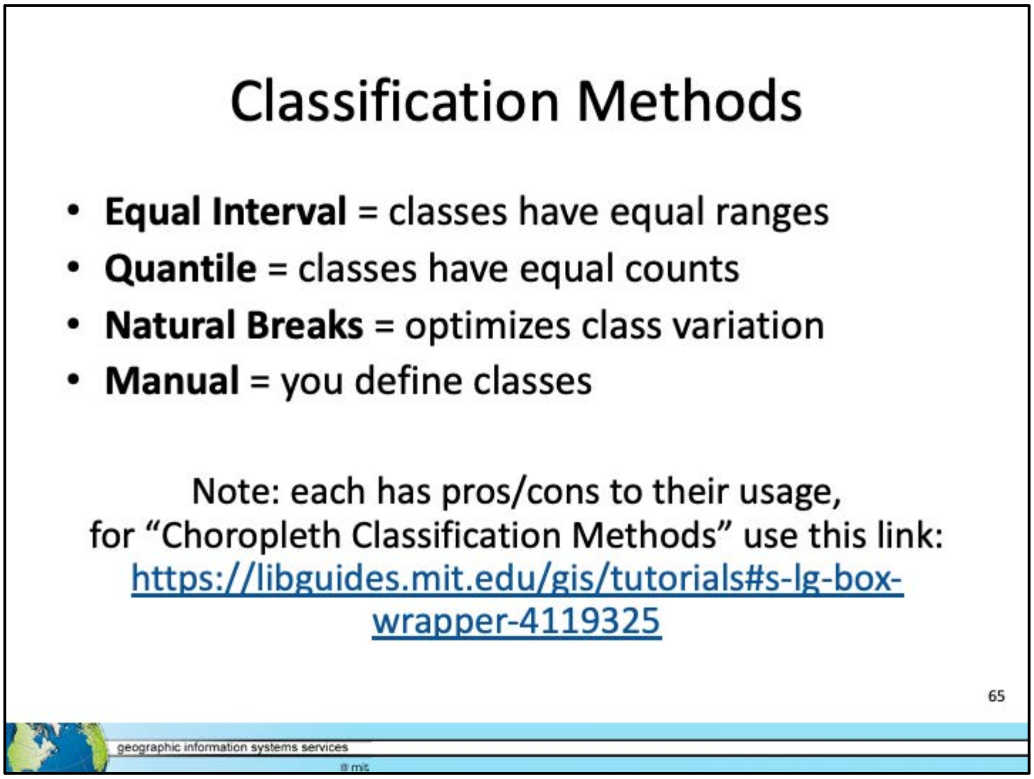
One of the most commonly used types of maps is a choropleth map. Choropleth maps use different shading and colors to display the quantity or value in defined areas. Choropleth maps are best used with polygon data so that it's easier to see color variations.

This example of a choropleth map uses shades of 2 different colors, orange and teal, to show spending per student by school districts. School district is the defined area and spending per student the quantity.



When designing a choropleth map, you have to make 2 basic choices:
-the number of classes you want your quantity value divided into
-the classification method for arranging the data into those classes

When choosing the number of classes, keep these points in mind:
-the more classes, the more variation you have . The human eye can't distinguish between large numbers of variations of the same color. It is best to have no more than 7 variations.
-the major types of classification method are Equal Intervals, Quantile, Natural Breaks and Defined Intervals.



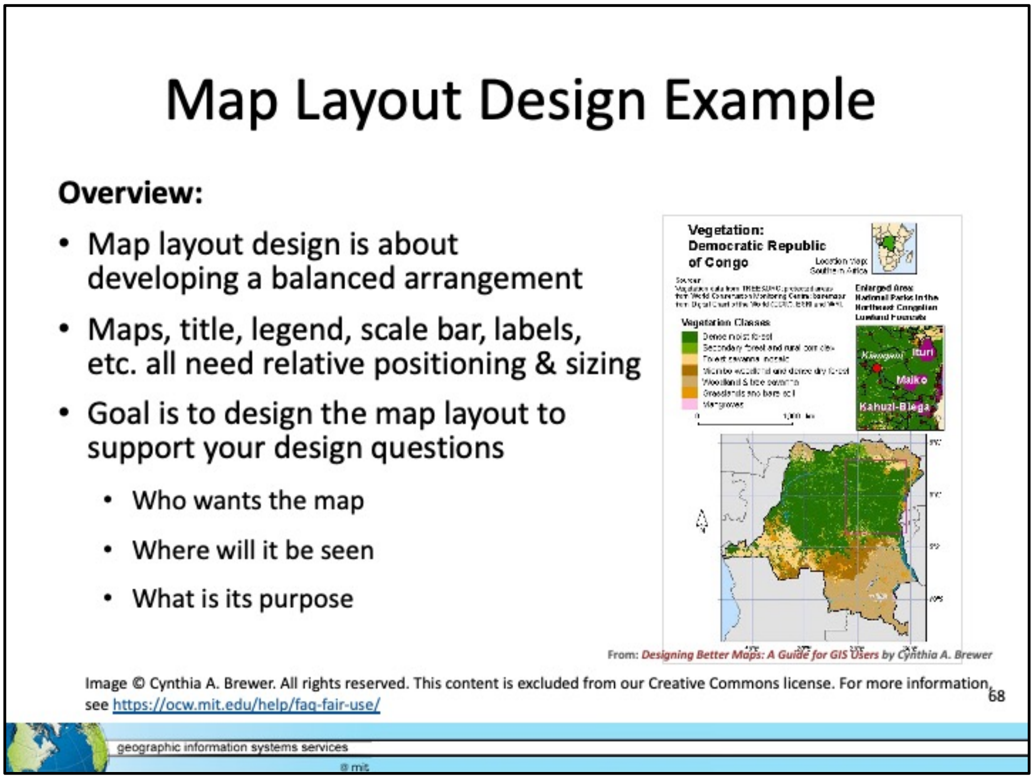
equal interval classification classes have equal ranges : ranges such as 1-5, 5-10, 10-15
quantile classification, classes have equal counts.: 5 items in each class
Natural breaks optimizes class variation: the algorithm figures out where the breaks should be
manual classification, the user sets the breaks based on prior knowledge of the data

It’s ok that you can’t see the exact numbers where the breaks are. The important thing is that these maps all use the same data (see histograms), but look different depending on the classification method used (see the amount of each color based on breaks in histograms).

Data source: SimplyAnalytics

Exercise 2

- Goal:
  - Learn how to symbolize different types of data
- Complete Exercise 2 for either QGIS or ArcGIS Pro.

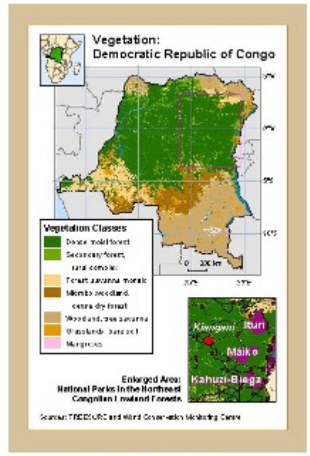


Once you have collected and symbolized your data, you are ready to create a layout for your map.
In this example, the eye tends to be drawn more to the legend and highlighted area than to the main map.

Map Layout Design Example

Tips:

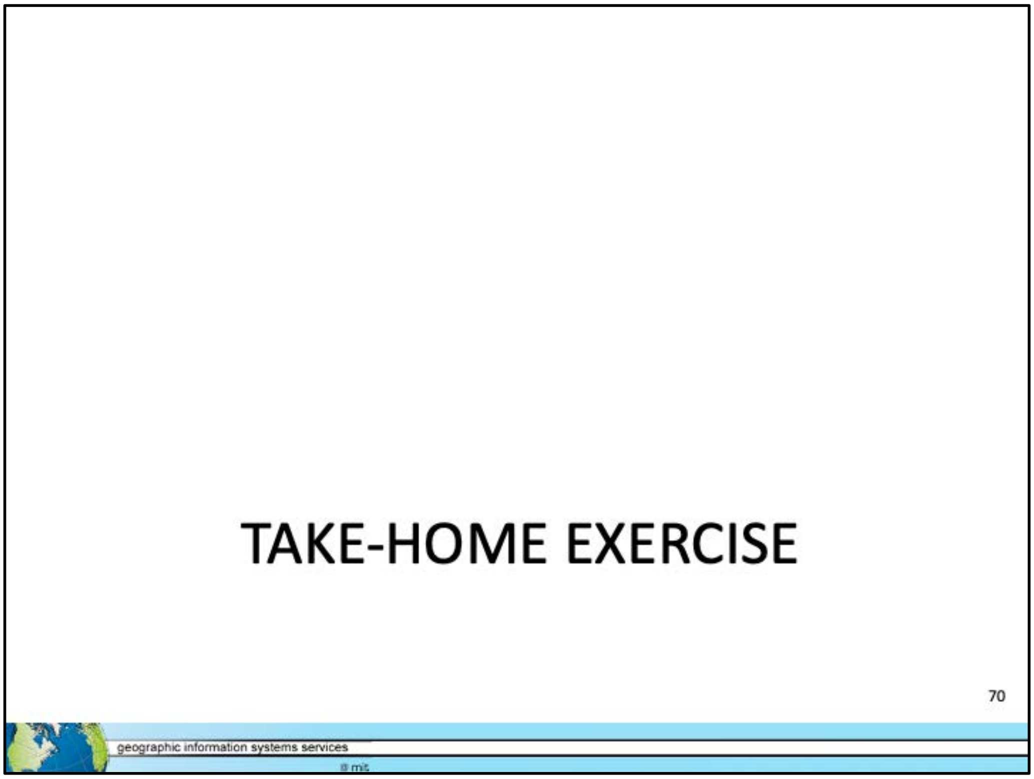
- Inset/locator maps are often placed in the top/bottom corners (e.g. continent view top left and zoomed view in bottom right).
- Main map often placed in center (usually largest & most detailed).
- Legend is tucked into the main map for easy comparison with the data.
- Scale bars and north arrows shouldn't be a distraction from the main map.
- Sources should run along the bottom.



From: *Designing Better Maps: A Guide for GIS Users* by Cynthia A. Brewer

Complete the take-home exercise to learn more.

In this rearranged map we have a much nicer visual flow, where the main map takes up the majority of the frame, and inset maps are tucked into the top and bottom corners for context.
Notice how the legend is not far from the main map for easy interpretation and there is room below the map for discussion and source credits.



Complete the take home exercise for either QGIS or ArcGIS Pro.

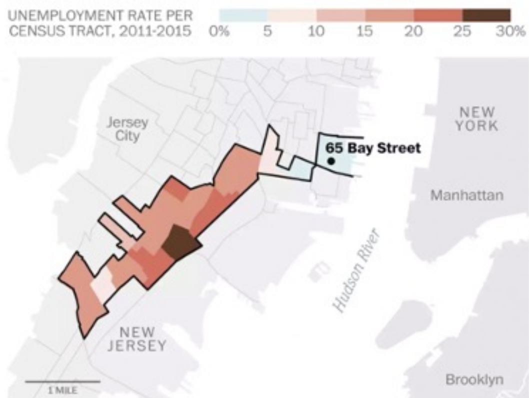
Exercise Overview

Query and use unemployment and transportation data to create a map that helps you decide where to build a mixed use facility.

1. Navigate the software interface
2. Find and add data, including basemaps
3. Access and explore attribute information
4. Symbolize data layers, for vector and raster
5. Select data by attributes and spatial location
6. Design a simple map for export

How 65 Bay St. was deemed part of a needy area

In the final map approved by state officials, 16 census tracts were linked together to connect the affluent Jersey City waterfront to impoverished and crime-ridden neighborhoods nearly four miles away. This allowed the project to qualify for low-interest loans through a U.S. visa program.



Source: Census Bureau

ANDREW TRAN AND GABRIEL FLORIT/THE WASHINGTON POST

Boburg, S. (2017, May 31). How Jared Kushner built a luxury skyscraper using loans meant for job-starved areas. *Washington Post*. Retrieved from

72

geographic information systems services

@ mit

Your exercise is based on data presented in a Washington Post article. The article was about Jared Kushner's use of maps to get federal funds meant for job starved areas, but many other developers have used maps for persuasion as well. This is a map from the article and it shows where the development was built. Data was left out to justify building in this location. Will you choose the same location? Where should the development really be built? If we have time, we'll look at some maps that you have created at the end of the workshop.

MIT OpenCourseWare
https://ocw.mit.edu

RES.STR-001 Geographic Information System (GIS) Tutorial
IAP 2022