



GIS Level 1: Introduction to GIS & Mapping

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

2

INTRODUCTION

3



Geographic Information System

“A system for
capturing, storing, checking, integrating,
manipulating, analyzing and displaying
spatial data”

4



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

A GIS system can do a lot of things.

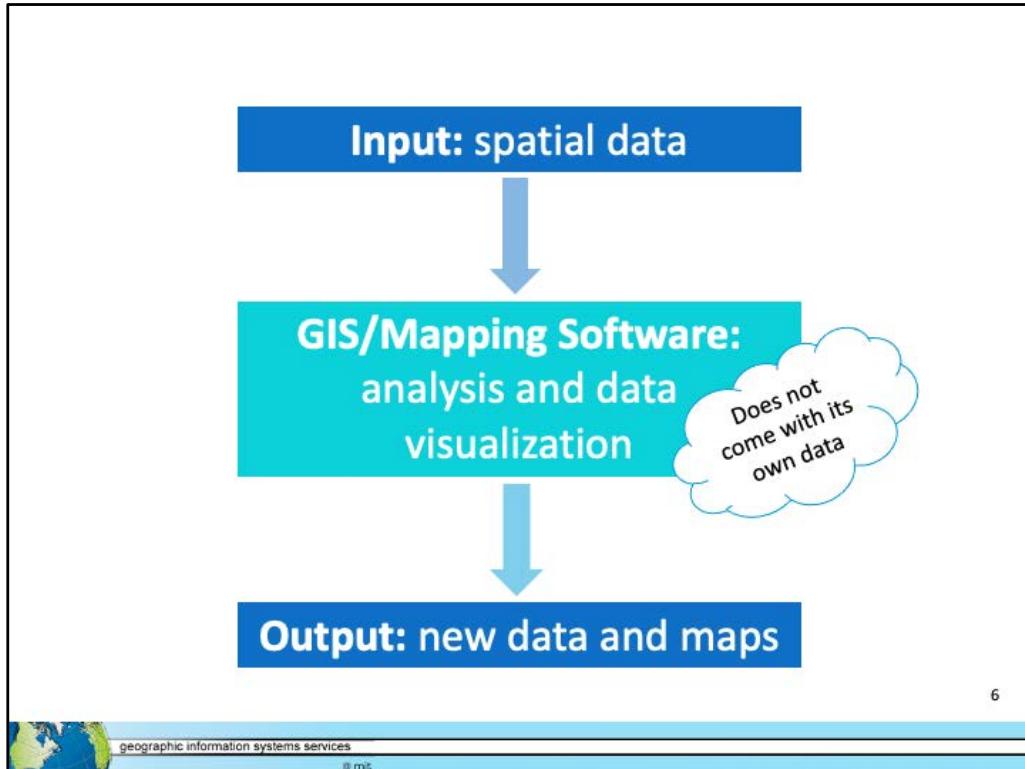
Geographic Information System

“A system for
capturing, storing, checking, integrating,
manipulating, analyzing and displaying
spatial data”

5



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



6

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



Image © source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

7



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

Theoretical Overview

GIS recreates real world spatial data

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

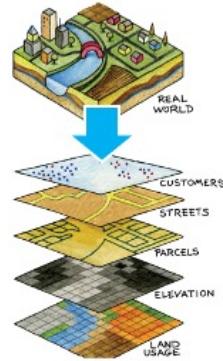


Image © source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

8

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

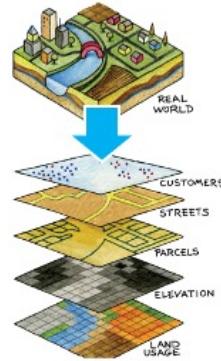


Image © source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

9

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



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

and overlaid for analysis

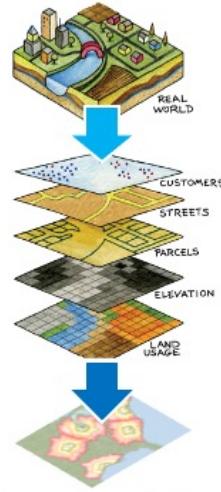


Image © source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

10

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



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

12



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 an 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.

Which desktop software should you use today?

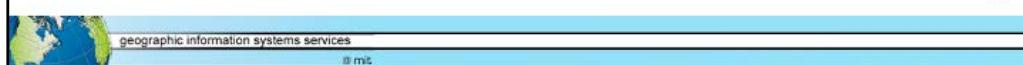
ArcGIS Pro (by ESRI)

- Commercial software (expensive to purchase)
- Only runs on Windows
- Larger program – can run slowly on some computers
- Full set of GIS functions and tools
- Integration with ArcGIS Online
- Fully developed training program (online modules, written tutorials, MOOCs)
- Comprehensive support (direct support from ESRI, documentation for every tool)

QGIS

- Free, open-source tool
- Runs on any operating system
- Smaller program that will not affect performance of your computer
- Many available tools, but lacking some for specific functions, such as network analysis (i.e. routing) and spatial statistics
- Basic tutorials by QGIS developers and users
- Tools can be developed by anyone so performance and documentation is inconsistent.
- Support via forums

13



- 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



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

View Imagery



City of Cambridge Aerial Photograph, April 2010

Image © source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

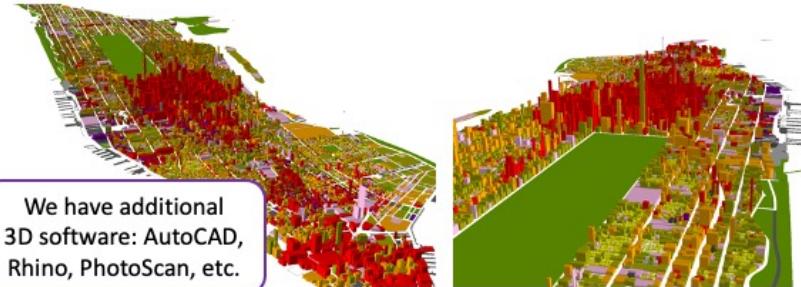
15



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

Create 3D models



We have additional
3D software: AutoCAD,
Rhino, PhotoScan, etc.

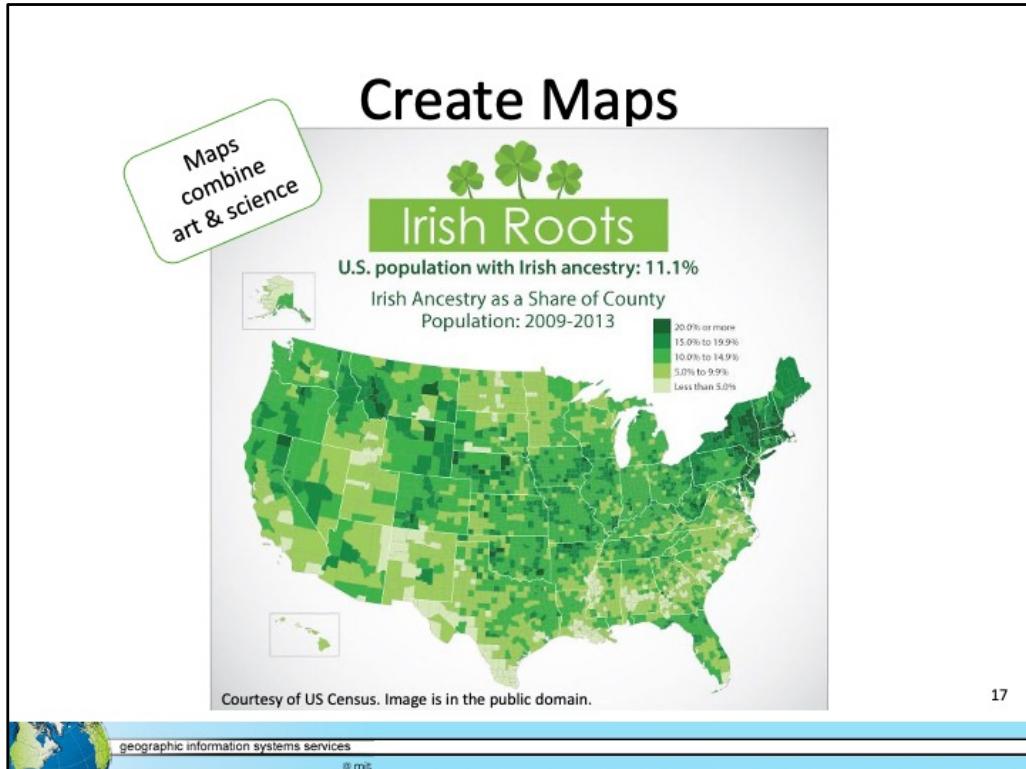
16



You can also make custom visualizations.

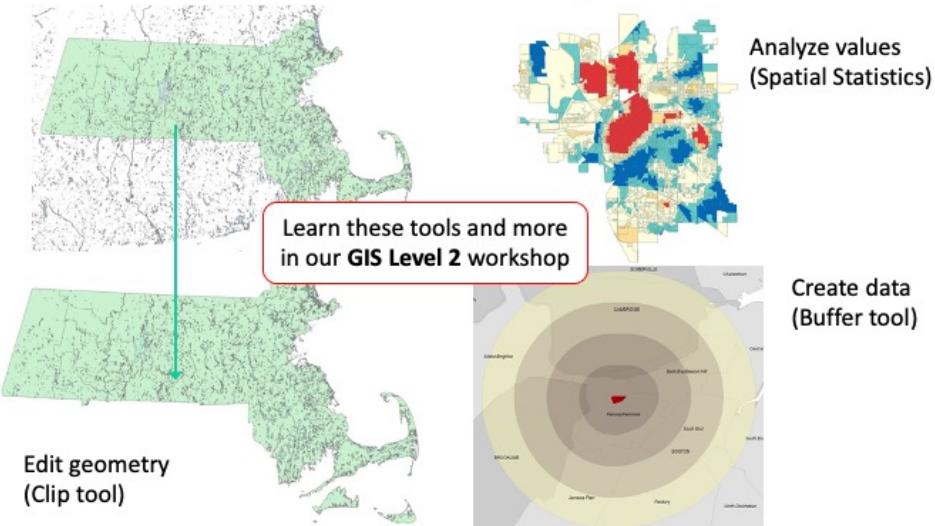
In this example, buildings in New York City have been color coded by land use and extruded vertically to represent height.

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

Conduct Analyses



Crime hotspots © Scott & Warmerdam. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

18

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

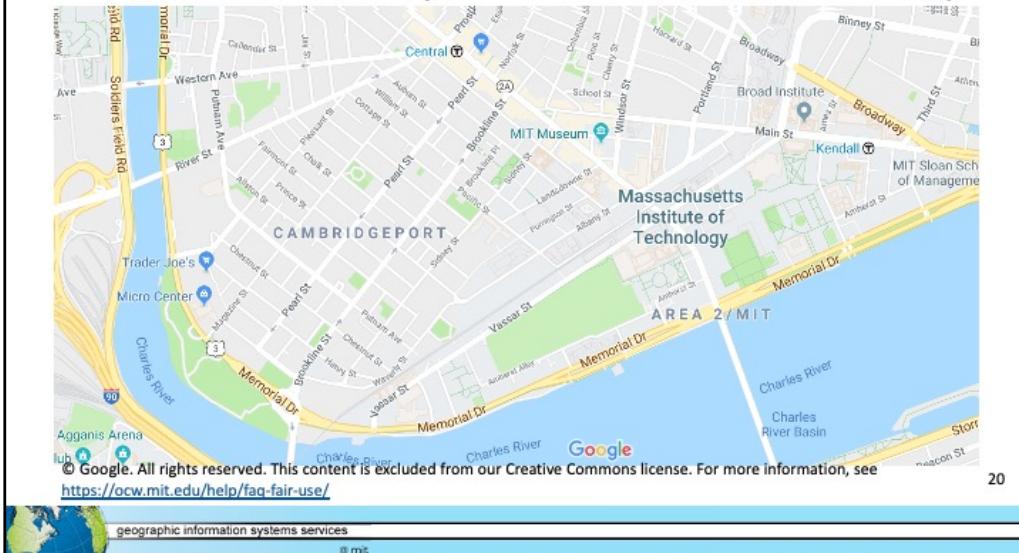
19



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

Understanding data ‘layers’

What individual data layers were used to create this map?



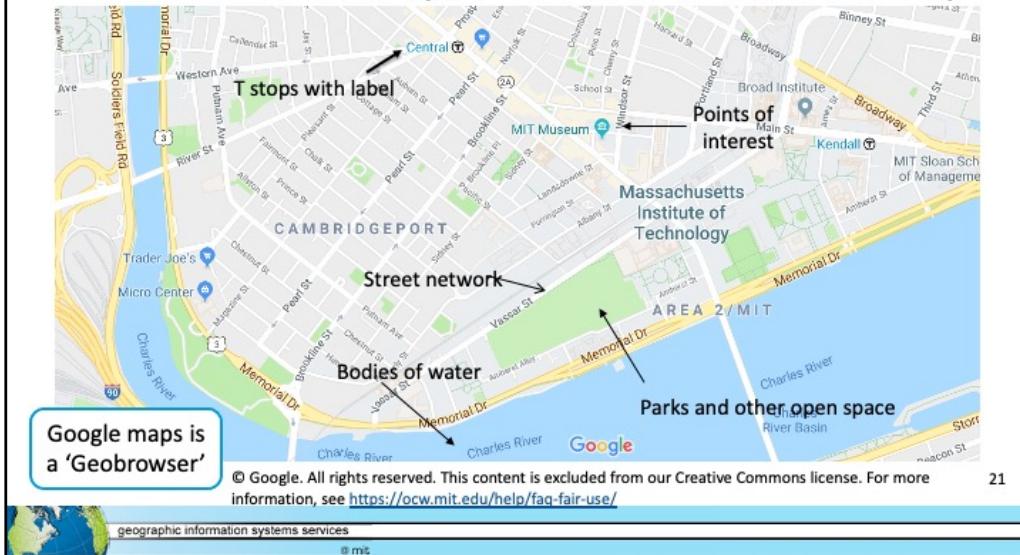
Google Maps is a type of GIS, really a web mapping tool or geobrowser, and its maps are created from multiple layers of spatial data. As shown previously, a layer is data about a specific set of similar features, such as the location of schools or bicycle paths.

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

Presenter should move on after a few responses come in.

Understanding data ‘layers’

What individual data layers were used to create this map?



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

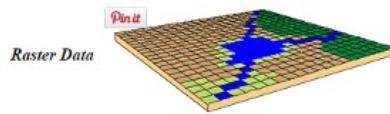
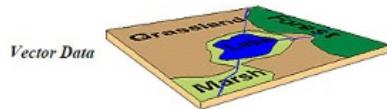
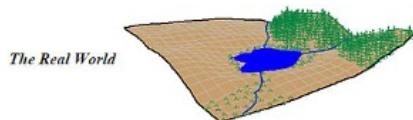
-When you create your own maps you can choose which data layers to include and how to visualize them to create a custom map.

MAPS & DATA: SPATIAL DATA TYPES

22



Geospatial Data Types



Geospatial or coordinate data can be represented in two different data formats:

Vector:

e.g. points, lines, and polygons

Raster:

e.g. row and column matrix

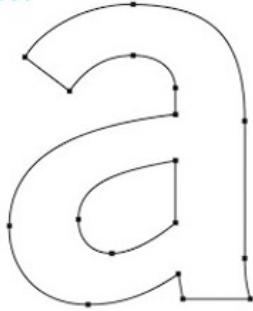
Images © University of Washington. All rights reserved.
This content is excluded from our Creative Commons
license. For more information, see
<https://ocw.mit.edu/help/faq-fair-use/>

23

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.

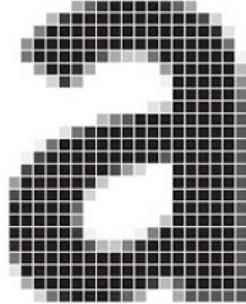
Data Types: Vector versus Raster

Vectors are composed of coordinates



VECTOR

Raster's are composed of pixels



RASTER

These are often used for variables with:
defined borders, e.g. manmade continuous surface, e.g. environmental

Image courtesy of [Zina Yonten](#). Used under CC BY-NC.

24



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.

Data Types: Vector examples

Points



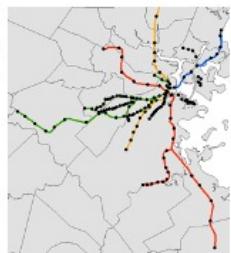
Lines



Polygons



(Combined)



25



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

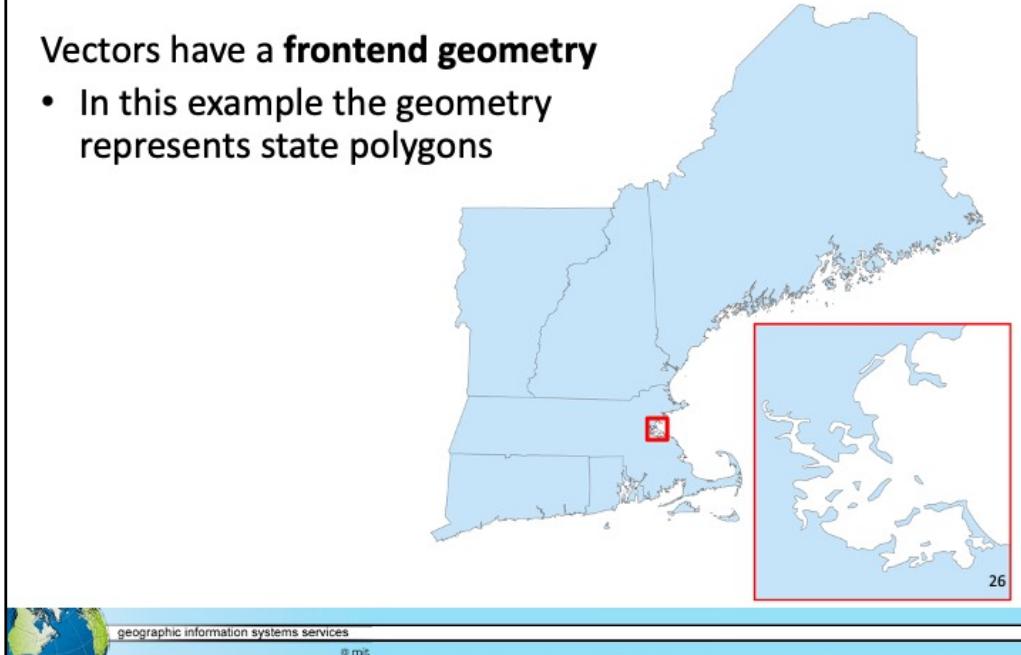
Question: Can anyone guess based on how these vector layers are color coded, also known as symbolized, what these datasets may represent?

Answer: MBTA Stops, Lines, and Towns, where the lines have been symbolized by color

Data Types: Vector mapping

Vectors have a **frontend geometry**

- In this example the geometry represents state polygons



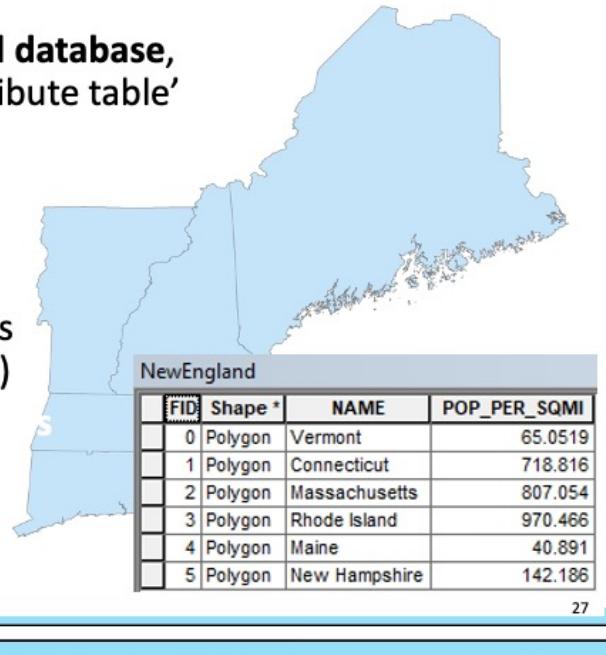
Geographic data include both a frontend geometry, meaning what you see on the screen in GIS software and a backend database.

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

Data Types: Vector mapping

Vectors have a **backend database**, normally called an 'attribute table'

- **rows** represent unique geometries (e.g. state polygons)
- **columns** represent a number of variables (theoretically infinite)



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.

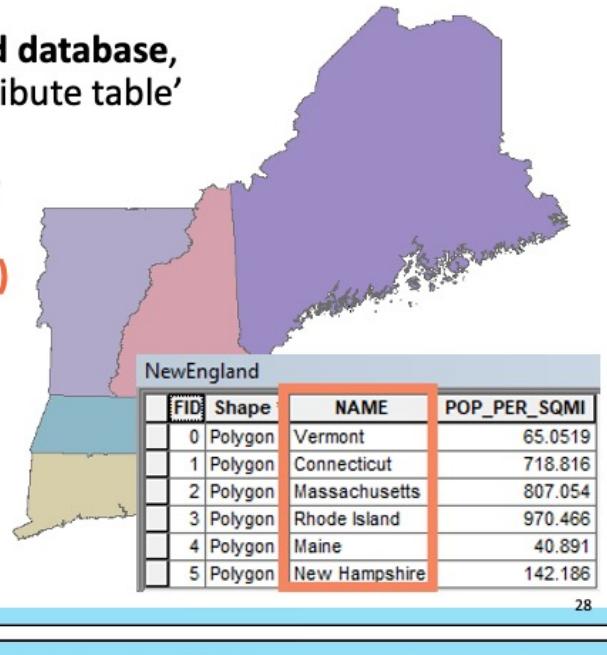
Each column is a different piece of information about that feature. In this example there is information about the state name and the population per square mile.

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

Data Types: Vector mapping

Vectors have a **backend database**, normally called an 'attribute table'

- Here each state is being symbolized by 'NAME' (qualitative variable)



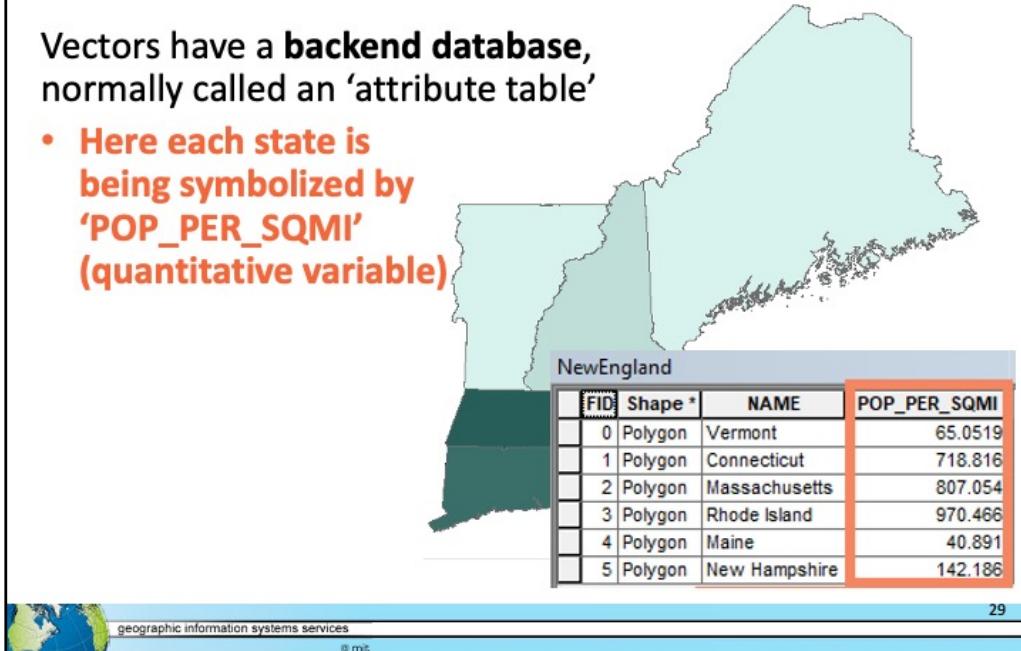
The map can be symbolized based on any column in the attribute table, meaning the color, size, shape, pattern, etc. of a feature can be changed to correspond to the data in a particular column.

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.

Data Types: Vector mapping

Vectors have a **backend database**, normally called an 'attribute table'

- Here each state is being symbolized by 'POP_PER_SQMI' (quantitative variable)



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.

Data Types: Vector file formats

- The shapefile is the most common vector file format.
- “A” shapefile is actually a collection of several different files with different extensions.

The diagram illustrates the components of a Shapefile. In the center, a blue box contains the text "Shapefile =" followed by a list of file extensions: ".shp .shx .sbx" on the first line and ".dbf .prj" on the second line. To the left of this central box, a red-bordered callout box contains the text: "When adding files to ArcGIS Pro, you will only see one file, not every extension." To the right, another red-bordered callout box contains the text: "Make sure to keep all files together when moving." At the bottom of the slide, there is a decorative footer bar featuring a globe icon, the text "geographic information systems services", and the number "30".

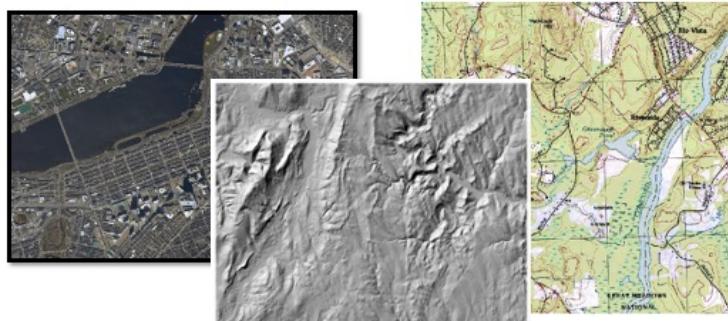
A shapefile is an open source format for vector data that can be opened in any GIS software.

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.

Data Types: Raster

Raster data includes aerial photographs, digital elevation models, and scanned maps.



(Remember these are constructed from pixels)

31

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

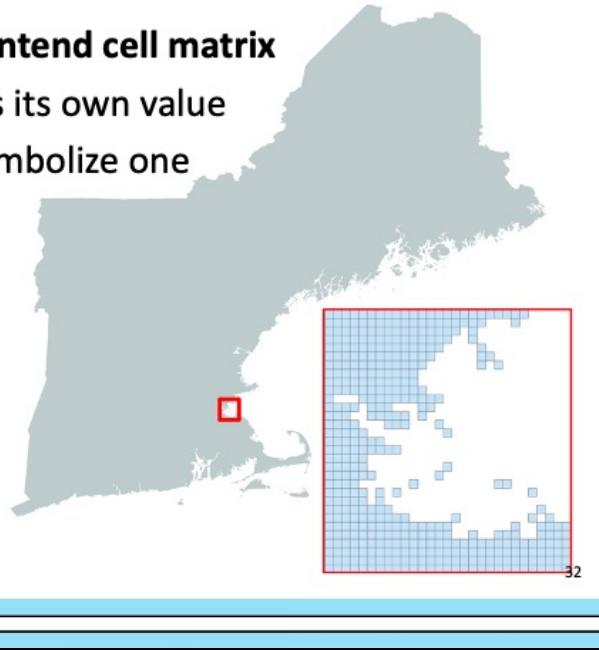
Additional Notes:

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

Data Types: Raster mapping

Raster data have a **frontend cell matrix**

- Where each cell has its own value
- A raster can only symbolize one variable at a time



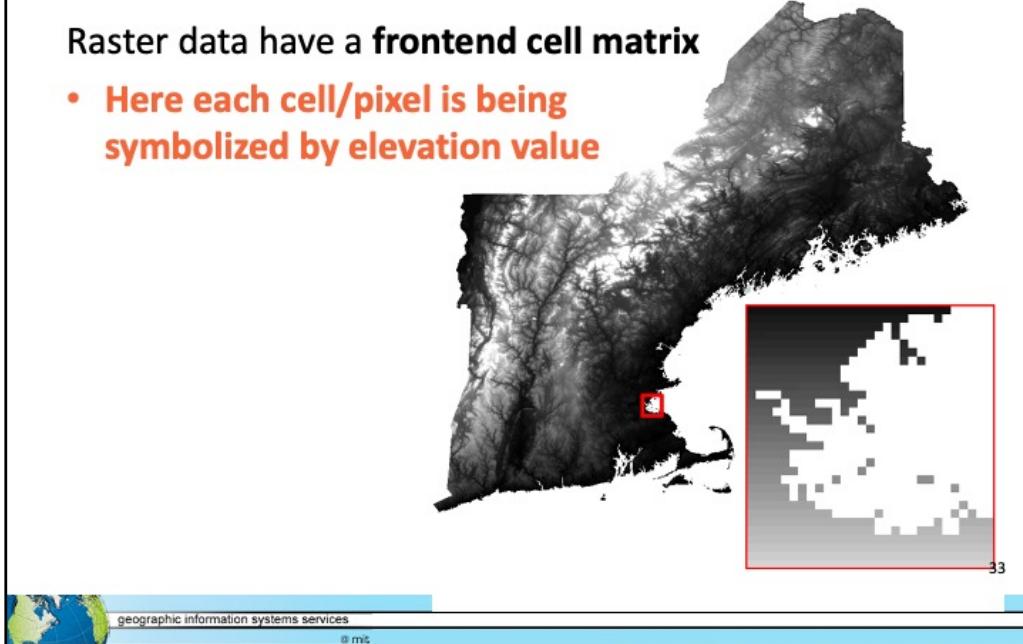
Raster data is a continuous cell matrix. Each cell or pixel is the same size and has its own value.

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

Data Types: Raster mapping

Raster data have a **frontend cell matrix**

- Here each cell/pixel is being symbolized by elevation value

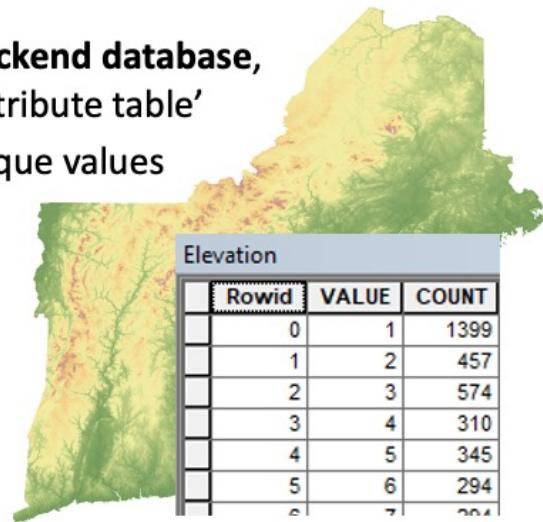


Here the map is color coded based on a quantitative (aka numerical) variable, elevation, where each unique pixel value is symbolized by a shade of grey stretching from black to white.

Data Types: Raster mapping

Raster data have a **backend database**, normally called an ‘attribute table’

- **rows** represent unique values (1m, 2m, 3m, etc.)
- **columns** have specific variables
 - 1) unique ‘ROW ID’
 - 2) unique ‘VALUE’
 - 3) ‘COUNT’ of pixels with that ‘VALUE’



34

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.

Data Types: Raster file formats

There are many different raster file extensions, including common image formats.

.tiff

.asc

.img

.jpg

Learn more about raster formats in this [ArcGIS Pro](#) documentation. QGIS supports similar formats.

Some formats may include a collection of files with different extensions, similar to a shapefile.

35



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.

Data Types: Tabular

Tabular data can be transformed into spatial data in two ways:



Created by Musavir Ahmed
Iron Nose Project

1. Joining

- Use a **shared unique identifier** (GEOID, name, etc.) to match up tabular data to the spatial data's attribute table.

2. Geocoding

- Use **lat/lon** coordinates in table to plot as points on map
- Use **addresses** to plot locations based on a street network

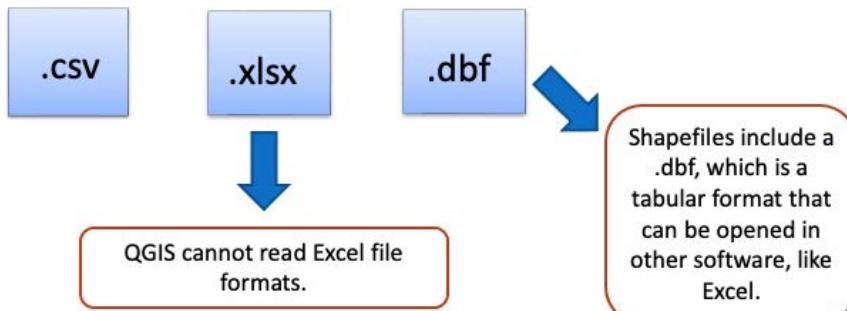
36



You can convert tabular data such as those in spreadsheets so long as 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.

Data Types: Tabular file formats

GIS software can read commonly used tabular formats in order to transform them into spatial data.

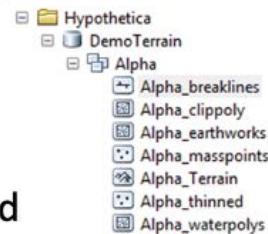


37

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.](#)



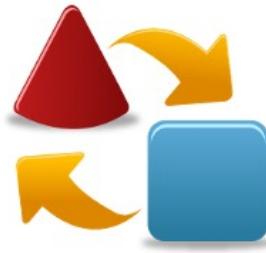
38

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.

Other data formats

GIS can import and convert data produced in other formats:

- KML / KMZ files (Google Earth)
- DXF / DWG (CAD)
- NetCDF (scientific data)
- LAS (Lidar)
- GPX (GPS units)
- Geojson



GIS software can export many formats:

- Adobe Illustrator
- KML
- CAD
- TIF
- JPG

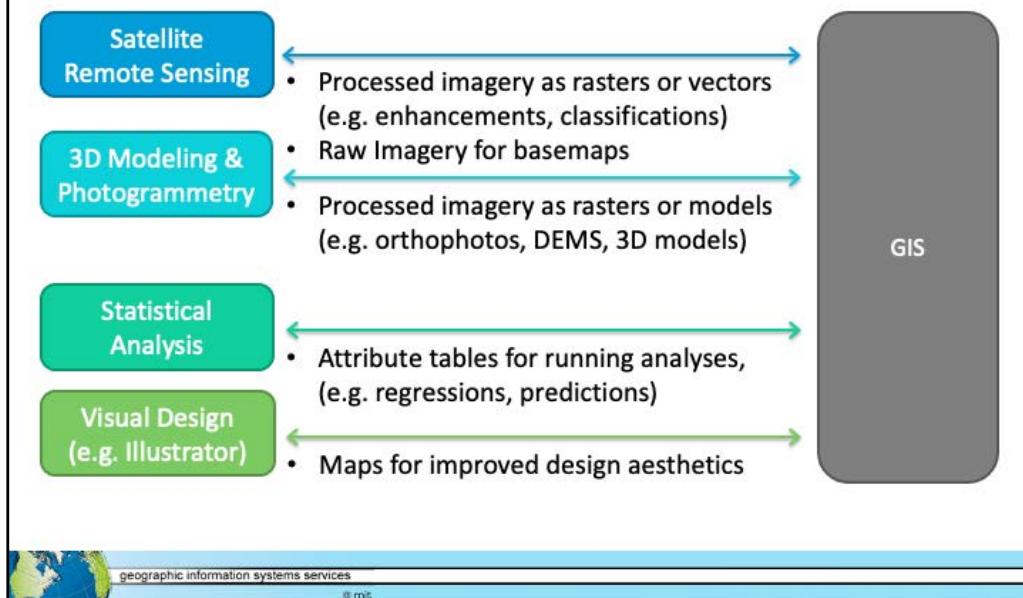
The GIS & Data Lab has
many types of data
visualization software.

39



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.

Common Associated Workflows



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.

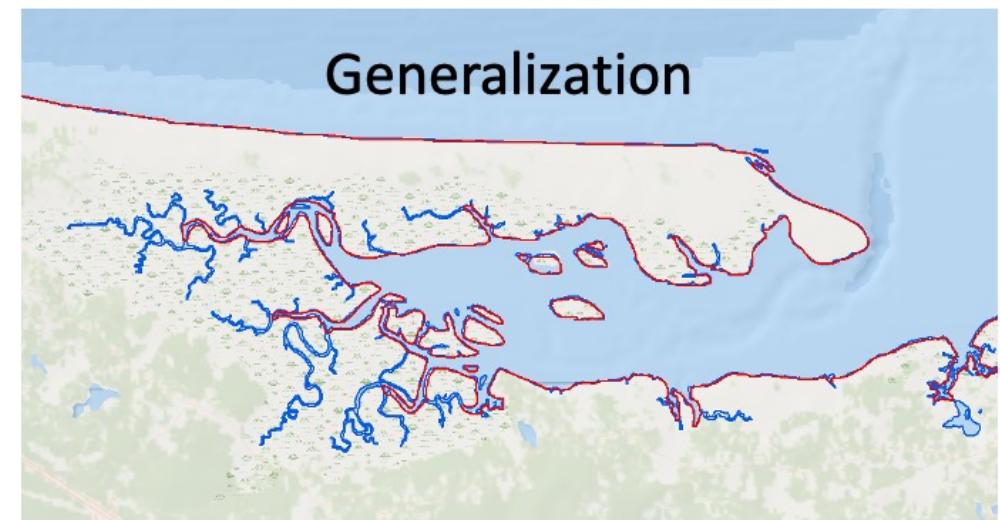
41



MAPS & DATA: CHARACTERISTICS OF SPATIAL DATA

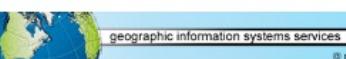
42





- The most detailed data available is not suitable for all purposes (or often a manageable file size)
- e.g. resolution of coastline data for this map is scale dependent
 - Red: county map
 - Blue: town map

43



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.

Abstraction

The process of reducing data from its complete state to what is necessary for use and presentation



Quiz: Which data symbology (pictured above) would you select for each of the following maps?

- Land use study of adjacent property
- Development map of the airport
- National map of airports

44

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 development 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)

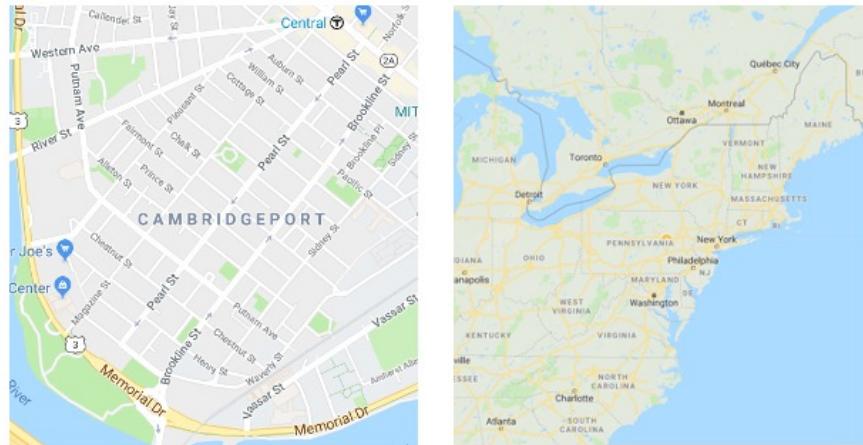
B - Because it just shows the airport as one symbol. If you used the symbols pictured in C or D, your map would be messy and difficult to read.



geographic information systems services

8 miles

Spatial Resolution/Scale



Suitable data geometry is dependent on scale:

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

© Google. All rights reserved. This content is excluded from our Creative Commons license. For more information, see

<https://ocw.mit.edu/help/faq-fair-use/>

geographic information systems services

45

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.

Temporal Resolution



Keep in mind temporal resolution when obtaining data

© Google. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

46

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.

Searching for Spatial Data

- Look in general GIS data repositories
- Search the internet
 - Include “gis”, or “data” in the search terms
 - Search by location and/or topic
- Search for country statistical agencies or open data sites (large cities often have their own open data portals as well)
- Contact GIS departments, universities, or researchers in your area of interest.
- Search for articles on your topic and look for the sources of the data.

★ Great slide to refer back to when starting a project

47



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.

Repositories and Websites

Libguides.mit.edu/gis

- Can also find by googling ‘MIT GIS’, first result
- Click on Find Data Tab for a list of resources, including an assembled links of common data sources per topic.

Find many more
on our website.

Geodata.mit.edu (Geoweb)

- Includes data licensed freely or restricted to MIT and other institutions, plus CDs and DVDs in the GIS lab.
- MIT instance is mainly historical-local or purchased data.

OpenStreetMap.org

- Crowd-sourced maps; content will vary by location
- Download as a shapefile via <http://www.geofabrik.de/>
- Best source to start for rural international data.

48



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.

MAPS & DATA: METADATA

49



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

What is Metadata?

Citation Information

Originator	MassGIS
Publication Date	20030301
Title	Massachusetts (Major Drainage Basins, 2003)
Geospatial Data Presentation Form	vector digital data
Publication Place	Massachusetts
Publisher	Massachusetts Office of Geographic and Environmental Information (MassGIS)
Online Linkage	Server=arrowsmith.mil.edu, Service=5150, Database=oracle

Description

Abstract This layer is a statewide digital datalayer of the 28 major drainage basins of Massachusetts as defined by the USGS Water Resources Division and the MA Water Resources Commission. It contains 865 arcs and 82,000 vertices. It is stored as a single statewide coverage. MassGIS created this data. MassGIS is maintaining this datalayer. In February 2002 the DEP GIS Group modified some linework within the USGS Worcester (#33) Quadrangle to synchronize the major basin divide with changes in the DEP Surface Water Supply Protection Area datalayers (Zone A, B, C) and Sub-basin tiles. In May 2002 the DEP GIS Group revised the N. Coastal/Pipewich basin boundary in Wrentham. In March 2003 linework was updated along the Westfield/Connecticut and N. Coastal/Pipewich boundaries.

Purpose This layer is intended for researchers, students, and policy makers for reference and mapping purposes, and may be used for basic applications such as viewing, querying, and map output production, or to provide a basemap to support graphical overlays and analysis with other spatial data. This layer will provide a basemap for layers related to real property and public utilities. More advanced user applications may focus on urban planning, public works, or utilities purposes (including defining boundaries, managing assets and facilities, integrating attribute databases with geographic features, spatial analysis, and presentation output.)

Time Period of Content

Calendar Date 20030301

geographic information systems services
© mits

Use metadata to learn how and why the data were created, access restrictions, columns in the attribute table, and much more!

50

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.

50

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



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



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

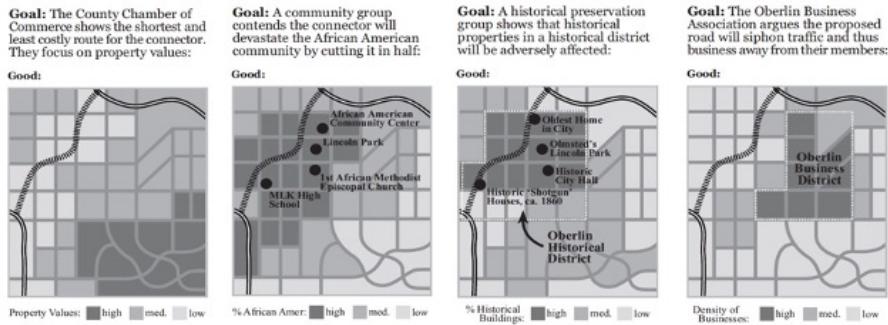


Art & science = design & analysis

Simplifications of reality = you can't show everything

Designed by people = there's a motive behind their creation

Making Great Maps



From: *Making Maps: A Visual Guide to Map Design for GIS* by John Krygier and Denis Wood

Example of how a map can be used to prove many different points.
However in the past only those in power had the software and data to do so.

Images © John Krygier and Denis Wood. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

54

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)

Three Key Questions

1. Who wants the map?

- e.g. experts (detailed), students (contextual), the community (interactive)

2. Where will it be seen?

- e.g. 8x11 paper (static small, room for main points)
- e.g. 30x40 poster board (static large, room for detail)
- e.g. web map (interactive, users control navigation of map)

3. What is its purpose?

- e.g. to show a variable through time (time series)
- e.g. to show change over time (change detection)
- e.g. to combine multiple variables into an index to pick best/worst (sustainability/risk/vulnerability mapping, site selection)

Each question deserves a well-thought answer before mapping

55



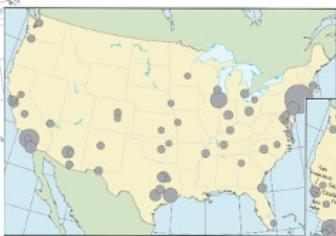
You also want to keep in mind some key questions, such as who wants the map and where it will be seen, both of which will affect the level of detail and whether or not the map should be interactive.

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.

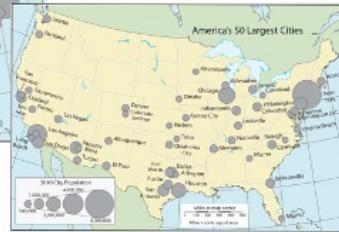
Map Design Process



Start with assembling the data from multiple sources



Next choose the data, analyses, & symbolization



Lastly insert the title, legend, north arrow, scale bar, & labels

56



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.)

Vector Symbolization

Feature Type	Visual Variable				Graphic Element		
	Size	Pattern Texture	Color Lightness	Color Saturation	Shape	Orientation	Color Hue
Point	Small Large	Low Coarse Fine	Low Light Dark	Pale Intense	Spring House Mine	Live Tree Dead Tree	Live Tree Dead Tree
Line	Medium Low	Medium High	Medium High	Medium Low	National Border Trail Section Line	Asphalt Road Concrete Road	National Border State Border
Area	High Medium Low	Medium High	Medium High	Medium High	Gravel	Orchard Sand	Land Field Crop Water

From: *Making Maps: A Visual Guide to Map Design for GIS* by John Krygier and Denis Wood - makingmaps.owu.edu

[See our tutorial](#) for additional
“Cartography Tips”.

[Colorbrewer](#) provides
accessible color options.

Images © John Krygier and Denis Wood. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

57

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 Symbolization

Symbolize your layer using one symbol



Stretch

Stretches values along a color ramp.



Discrete

Groups data based on a selected number of colors and applies a color scheme.



Classify

Assigns a color for each group of values.



Unique Values

Assigns a color for each value.



Vector Field

Displays values as vector symbols.

58



geographic information systems services

© mts

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.

Choosing Color Tips

Match the type of data to the type of color scheme:

- **Qualitative (categories)**

DISCRETE CLASSES
Qualitative

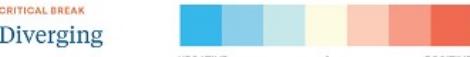


- **Quantitative (numbers)**

SCALAR VALUES
Sequential



CRITICAL BREAK
Diverging



Images © Morphocode. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>



When it comes to choosing color, you want to be mindful of the 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.

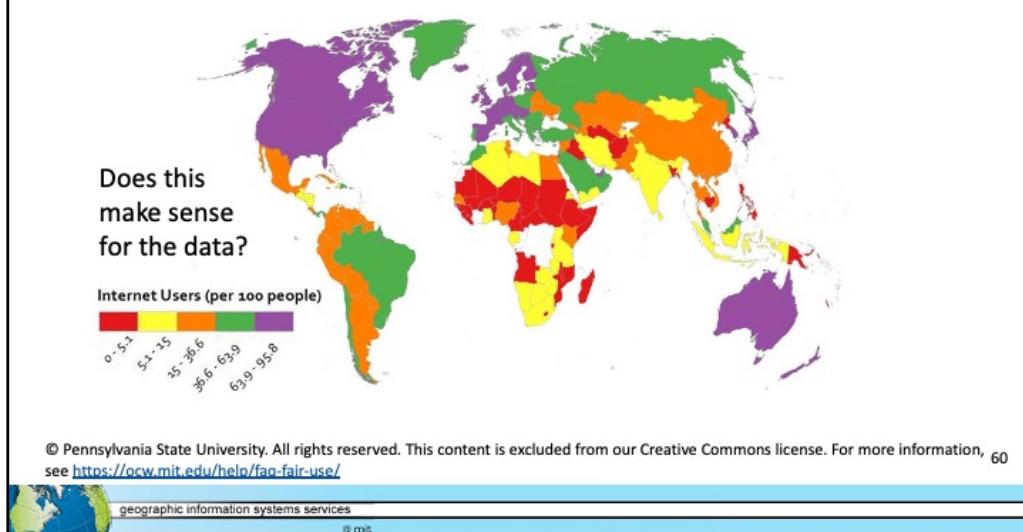
Examples of each would be:

Example of qualitative/categorical: land use

Example of sequential numbers: population

Example of diverging numbers: weather/political (e.g. red, blue, neutral)

Qualitative Color Example



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.

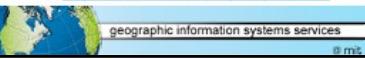
Sequential Color Example

Does this
make sense
for the data?

Internet Users (per 100 people)

Category	Range (per 100 people)
0-5.2	0-5.2
5.3-15.6	5.3-15.6
15.7-36.6	15.7-36.6
36.7-63.9	36.7-63.9
64.0-93.8	64.0-93.8
93.9+	93.9+

© Pennsylvania State University. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/> 61

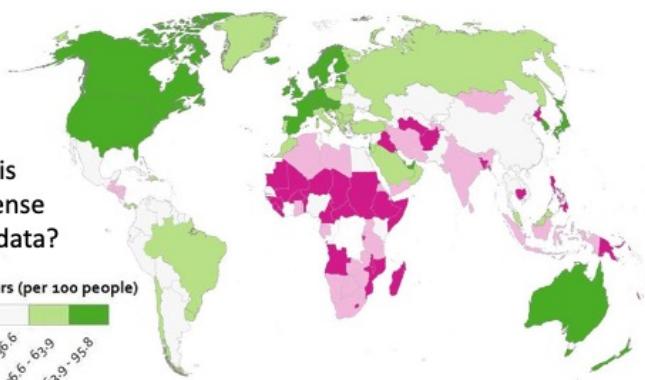
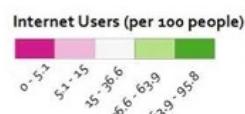


Question: Can you tell where the highest internet users are?

Answer: Yes, this has an ascending trend which is reflected in the darker color indicating greater intensity.

Diverging Color Example

Does this
make sense
for the data?



© Pennsylvania State University. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/> 62

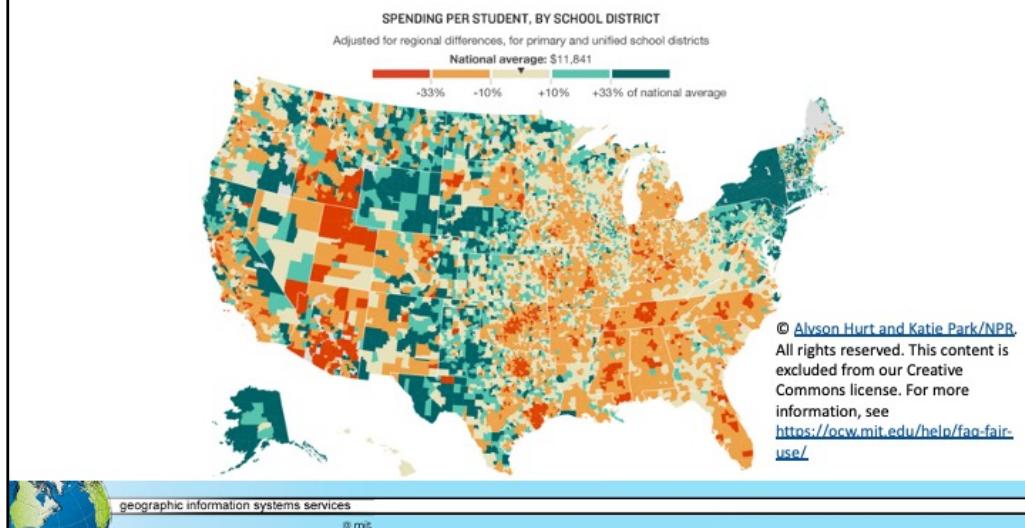


Question: Can you tell where the highest internet users are?

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).

Commonly used map type: Choropleth

These use different shading and coloring to display the quantity or value in defined areas.



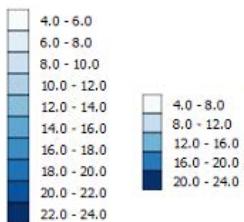
One of the most commonly used types of maps is a choropleth map. Choropleth maps use different shading and colors to display the quantity or value in defined areas. Choropleth maps are best used with polygon data so that it's easier to see color variations.

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.

Choropleth map choices

1. Number of Classes

- Aggregates data for display
- More classes = more variation (best to have no more than 7)



2. Classification Method

- Data classification is how data is arranged into separate classes.
- Major types
 - Equal Intervals
 - Quantile (Equal Count)
 - Natural Breaks
 - Defined Intervals

64



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.

Classification Methods

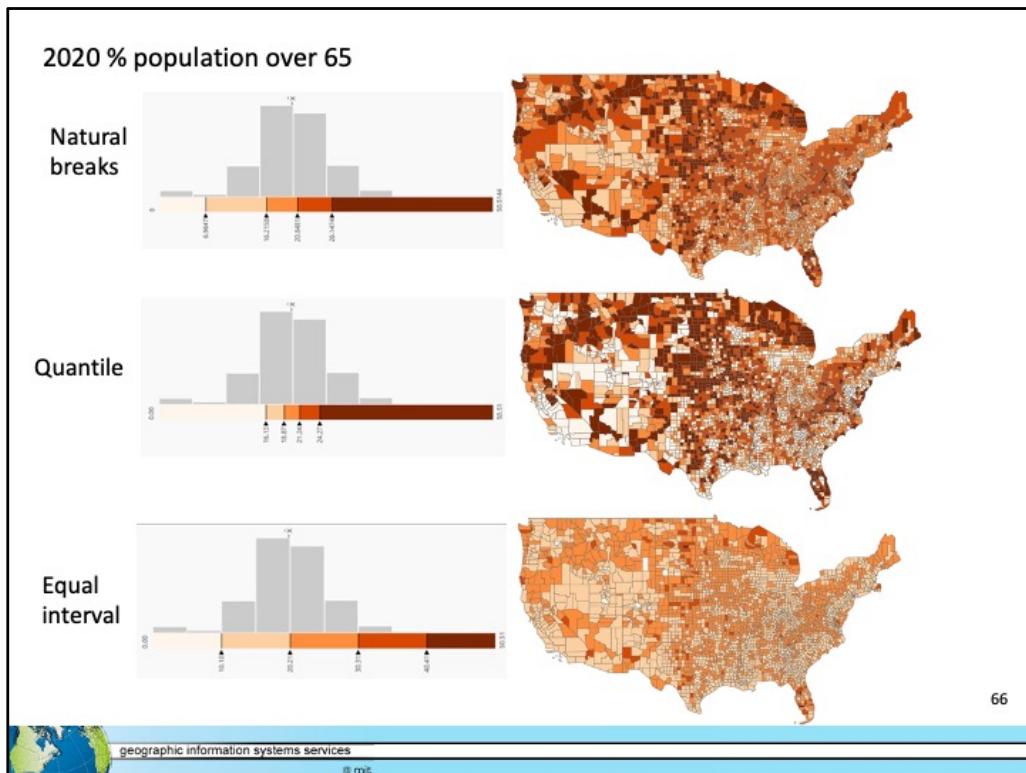
- **Equal Interval** = classes have equal ranges
- **Quantile** = classes have equal counts
- **Natural Breaks** = optimizes class variation
- **Manual** = you define classes

Note: each has pros/cons to their usage,
for “Choropleth Classification Methods” use this link:
<https://libguides.mit.edu/gis/tutorials#s-lg-box-wrapper-4119325>

65



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.

67



Map Layout Design Example

Overview:

- Map layout design is about developing a balanced arrangement
- Maps, title, legend, scale bar, labels, etc. all need relative positioning & sizing
- Goal is to design the map layout to support your design questions
 - Who wants the map
 - Where will it be seen
 - What is its purpose

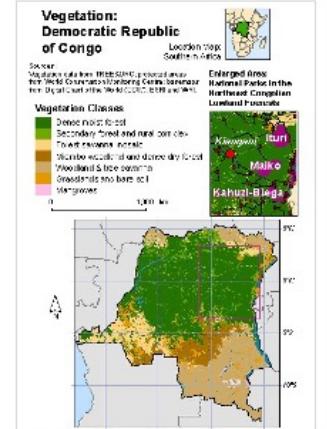


Image © Cynthia A. Brewer. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

Once you have collected and symbolized your data, you are ready to create a layout for your map.

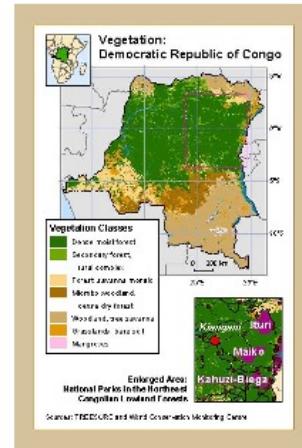
In this example, the eye tends to be drawn more to the legend and highlighted area than to the main map.

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.

Complete the take-home exercise to learn more.



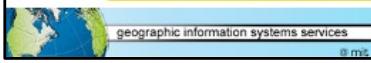
From: *Designing Better Maps: A Guide for GIS Users* by Cynthia A. Brewer

Image © Cynthia A. Brewer. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>

69

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.



© mits

TAKE-HOME EXERCISE

70



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

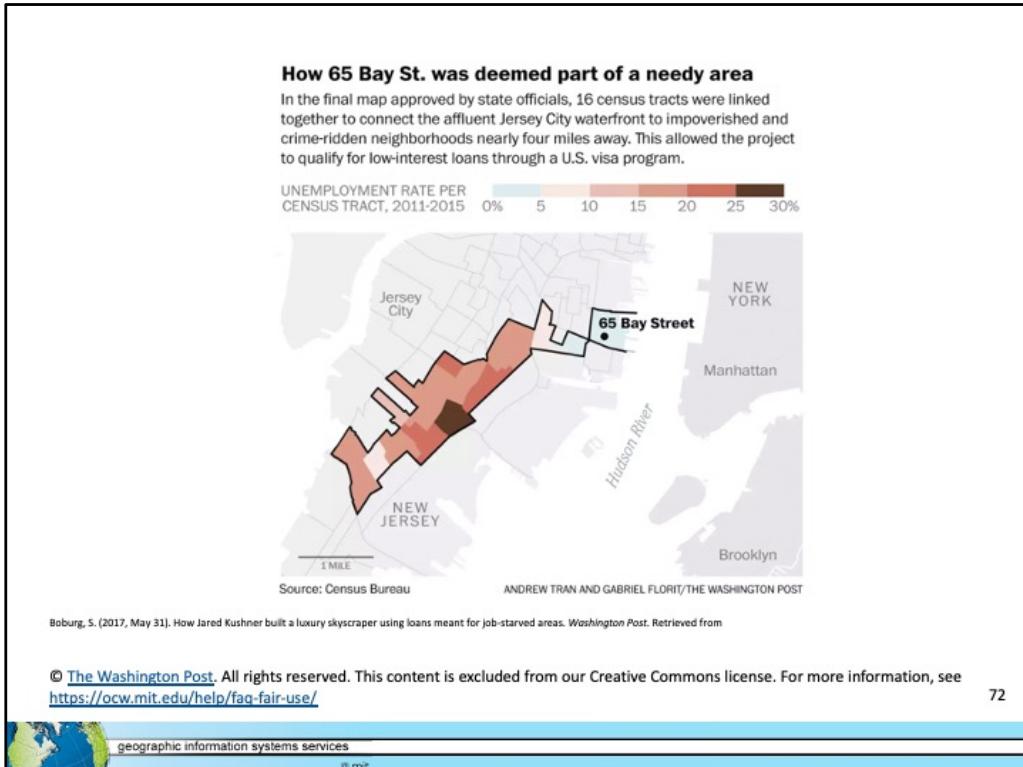
Exercise Overview

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

71





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

For information about citing these materials or our Terms of Use, visit:

<https://ocw.mit.edu/terms>

73

