GIS Level 1: Introduction to GIS & Mapping
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
Geographic Information System

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

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

GIS/Mapping Software: analysis and data visualization

Output: new data and maps

Does not come with its own 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/
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/
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/
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/
SOFTWARE
# Types of GIS & Mapping Software

<table>
<thead>
<tr>
<th>Type</th>
<th>Analysis Power</th>
<th>Example(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geobrowser</strong></td>
<td>Weak</td>
<td>Google Maps, Google Earth, Apple Maps, Waze, etc.</td>
</tr>
<tr>
<td></td>
<td>(mainly only to display data)</td>
<td></td>
</tr>
<tr>
<td><strong>Web-based</strong></td>
<td><strong>Medium</strong></td>
<td>Carto, ArcGIS Online, Mapbox, Google MyMaps, etc.</td>
</tr>
<tr>
<td></td>
<td>(able to upload additional data, customize display,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and perform basic analyses)</td>
<td></td>
</tr>
<tr>
<td><strong>Desktop</strong></td>
<td><strong>Strong</strong></td>
<td>ArcGIS Pro, QGIS</td>
</tr>
<tr>
<td></td>
<td>(installed locally, provides full control of map creation, and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>perform advanced analyses)</td>
<td></td>
</tr>
</tbody>
</table>
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
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/
Create 3D models

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

Maps combine art & science

Irish Roots
U.S. population with Irish ancestry: 11.1%
Irish Ancestry as a Share of County Population: 2009-2013

20.0% or more
15.0% to 19.9%
10.0% to 14.9%
5.0% to 9.9%
Less than 5.0%

Courtesy of US Census. Image is in the public domain.
Conduct Analyses

Analyze values (Spatial Statistics)

Create data (Buffer tool)

Learn these tools and more in our GIS Level 2 workshop

Edit geometry (Clip tool)

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/
UNDERSTANDING MAPS & DATA

You may have been looking at geospatial data for a long time
Understanding data ‘layers’

What individual data layers were used to create this map?
Understanding data ‘layers’

What individual data layers were used to create this map?

- Street network
- Parks and other open space
- T stops with label
- Bodies of water
- Points of interest

Google maps is a ‘Geobrowser’

© 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/
MAPS & DATA:
SPATIAL DATA TYPES
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
Data Types: Vector versus Raster

Vectors are composed of coordinates

Raster's are composed of pixels

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.
Data Types: Vector examples

Points

Lines

Polygons

(Combined)
Data Types: Vector mapping

Vectors have a **frontend geometry**

- In this example the geometry represents state polygons
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)
Data Types: Vector mapping

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

- Here each state is being symbolized by ‘NAME’ (qualitative variable)
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)
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.

Shapefile = .shp .shx .sbx .dbf .prj

When adding files to ArcGIS Pro, you will only see one file, not every extension.

Make sure to keep all files together when moving.
Data Types: Raster

Raster data includes aerial photographs, digital elevation models, and scanned maps. (Remember these are constructed from pixels)
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
Data Types: Raster mapping

Raster data have a **frontend cell matrix**

- Here each cell/pixel is being symbolized by elevation value
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’
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](https://pro.arcgis.com) documentation. QGIS supports similar formats.

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

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

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
Data Types: Tabular file formats

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

- .csv
- .xlsx
- .dbf

Shapefiles include a .dbf, which is a tabular format that can be opened in other software, like Excel.

QGIS cannot read Excel file formats.
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.
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.
Common Associated Workflows

Satellite Remote Sensing
- Processed imagery as rasters or vectors (e.g. enhancements, classifications)
- Raw Imagery for basemaps

3D Modeling & Photogrammetry
- Processed imagery as rasters or models (e.g. orthophotos, DEMS, 3D models)

Statistical Analysis
- Attribute tables for running analyses, (e.g. regressions, predictions)

Visual Design (e.g. Illustrator)
- Maps for improved design aesthetics
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.
MAPS & DATA:
CHARACTERISTICS OF SPATIAL DATA
Generalization

- 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
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
Spatial Resolution/Scale

Suitable data geometry is dependent on scale:

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

**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/](http://www.geofabrik.de/)
- Best source to start for rural international data.
MAPS & DATA: METADATA
What is Metadata?

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

<table>
<thead>
<tr>
<th>Citation Information</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Originator</strong></td>
<td>MassGIS</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>20030301</td>
</tr>
<tr>
<td><strong>Title</strong></td>
<td>Massachusetts (Major Drainage Basins, 2003)</td>
</tr>
<tr>
<td><strong>Geospatial Data Presentation Form</strong></td>
<td>vector digital data</td>
</tr>
<tr>
<td><strong>Publication Place</strong></td>
<td>Massachusetts</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>Massachusetts Office of Geographic and Environmental Information (MassGIS)</td>
</tr>
<tr>
<td><strong>Online Linkage</strong></td>
<td>Server=arrowsmith.mit.edu; Service=5150; Database=oracle</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abstract</strong></td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time Period of Content</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Calendar Date</strong></td>
<td>20030301</td>
</tr>
</tbody>
</table>
Metadata Examples


2. GeoWeb: geodata.mit.edu/catalog/mit-w37ehgh6nvl4w

3. City of Boston: https://data.boston.gov/dataset/traffic-signals
MAKING GREAT MAPS:
DATA VISUALIZATION PRINCIPLES
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
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/
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
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
# Vector Symbolization

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Visual Variable</th>
<th>Graphic Element</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Size</td>
<td>Shape</td>
</tr>
<tr>
<td></td>
<td>Pattern Texture</td>
<td>Orientation</td>
</tr>
<tr>
<td></td>
<td>Color Lightness</td>
<td>Color Hue</td>
</tr>
<tr>
<td></td>
<td>Color Saturation</td>
<td></td>
</tr>
<tr>
<td>Point</td>
<td>Small&lt;br&gt;Large</td>
<td>Spring&lt;br&gt;Live Tree&lt;br&gt;Live Tree</td>
</tr>
<tr>
<td></td>
<td>Low&lt;br&gt;High</td>
<td>House&lt;br&gt;Dead Tree&lt;br&gt;Dead Tree</td>
</tr>
<tr>
<td></td>
<td>Coarse&lt;br&gt;Fine</td>
<td>Mine&lt;br&gt;Mine</td>
</tr>
<tr>
<td></td>
<td>Low&lt;br&gt;High</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Light&lt;br&gt;Dark</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low&lt;br&gt;High</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pale&lt;br&gt;Intense</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low&lt;br&gt;High</td>
<td></td>
</tr>
<tr>
<td>Line</td>
<td>Medium</td>
<td>National Border&lt;br&gt;State Border</td>
</tr>
<tr>
<td></td>
<td>Low&lt;br&gt;High</td>
<td>Asphalt Road&lt;br&gt;Concrete Road</td>
</tr>
<tr>
<td></td>
<td>Low&lt;br&gt;High</td>
<td>Trail&lt;br&gt;Section Line</td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td>High&lt;br&gt;Medium</td>
<td>Land&lt;br&gt;Field Crop</td>
</tr>
<tr>
<td></td>
<td>Medium&lt;br&gt;Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium&lt;br&gt;High</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium&lt;br&gt;Low</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium&lt;br&gt;High</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium&lt;br&gt;Low</td>
<td></td>
</tr>
</tbody>
</table>

*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/](https://ocw.mit.edu/help/faq-fair-use/)
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.
Choosing Color Tips

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

- **Qualitative** (categories)
  - Discrete Classes
    - Qualitative
  - Sequential
  - Diverging

- **Quantitative** (numbers)

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/
Qualitative Color Example

Does this make sense for the data?

Internet Users (per 100 people)

© 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/
Sequential Color Example

Does this make sense for the data?

Internet Users (per 100 people)

© 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/
Diverging Color Example

Does this make sense for the data?

Internet Users (per 100 people)

0-5.1
5.1-15
15.1-36.6
36.6-69.9
69.9-99.8

© 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/
Commonly used map type: Choropleth

These use different shading and coloring to display the quantity or value in defined areas.
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
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
2020 % population over 65

Natural breaks

Quantile

Equal interval
Exercise 2

• Goal:
  – Learn how to symbolize different types of data

• Complete Exercise 2 for either QGIS or ArcGIS Pro.
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/
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.

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/
TAKE-HOME EXERCISE
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

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