All About Spatial Data

Find it, Manage it, Use it

Workshop Outline

- Define spatial data
- Spatial data formats
- Data Sources
- Metadata
- Evaluating Spatial Data
- Data Management
- Activities: define spatial data, find and open data, read metadata, interactive polls

Why Spatial Data?

- ## It's everywhere.
  - Spatial data is all around us, especially because of social media and GPS enabled devices. It's not just in the form of paper maps now.

- ## It's confusing.
  - It can come in many different formats and file types.

- ## It's not just a GIS thing.
  - Many types of software and programming languages can be used to view and analyze spatial data.

What words or phrases do you associate with spatial data?

What is spatial data?

- Information about the locations and shapes of geographic features and the relationships between them, usually stored as coordinates and topology.

- Any data that can be mapped.
 - This is the definition we'll use for the workshop.

  http://support.esri.com/en/knowledgebase/GISDictionary/term/spatial%20data

Examples of Spatial Data

- An Excel file with addresses

- A picture with coordinates

- A vector data layer (a line representing a river, polygons representing census tracts, etc.)

- An image of a map

Spatial Data Formats

- Vector:
  - Can usually be opened in a web map or GIS software with no additional processing.
  - Represent information using points, line, and polygons.
  - .shp/shapefile (saved as .zip for uploading), .gpx, .kml, kmz, .geojson, .osm, .bz2
- Tabular:
  - Will need to be "georeferenced" or "geocoded" to translate coordinates or addresses into vector shapes (points, lines, etc.).
  - Can be joined to vector data if there is a common ID.
  - .csv, .txt, .xls, .xlsx, .tab, .sql, .ods
- Raster/Image:
  - Pixelated data that can be added to a map, but cannot always be edited.
  - May need to be georeferenced.
  - Geotiff, .jpg (and other image formats)
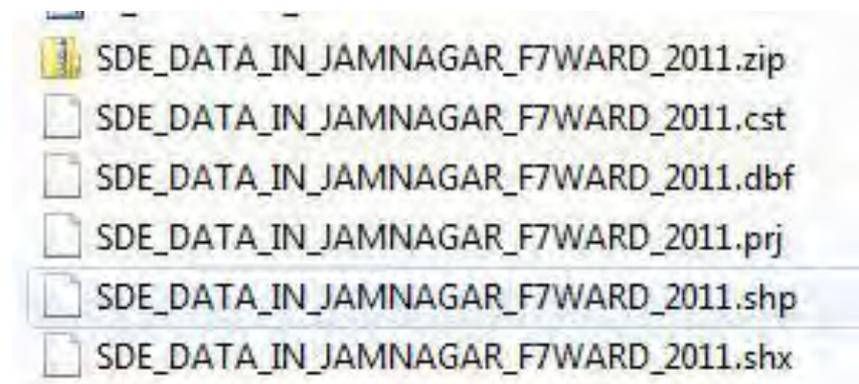
Vector: Points, Lines and Polygons
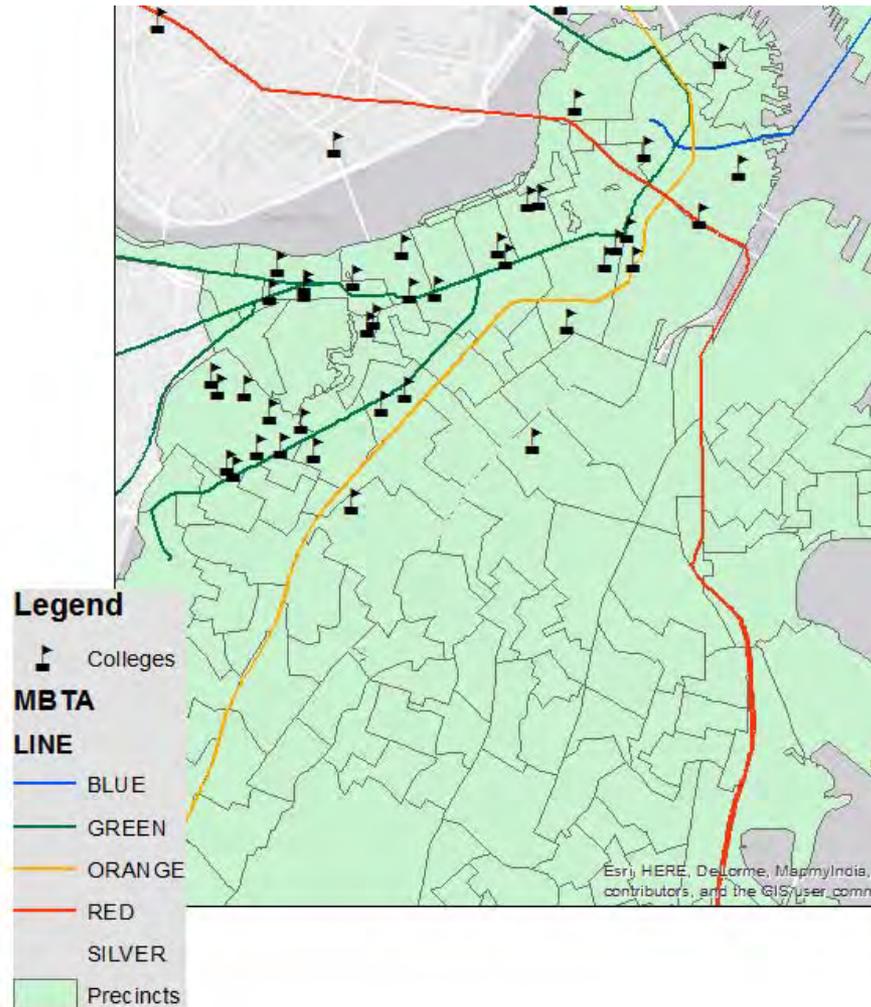
Attribute Table
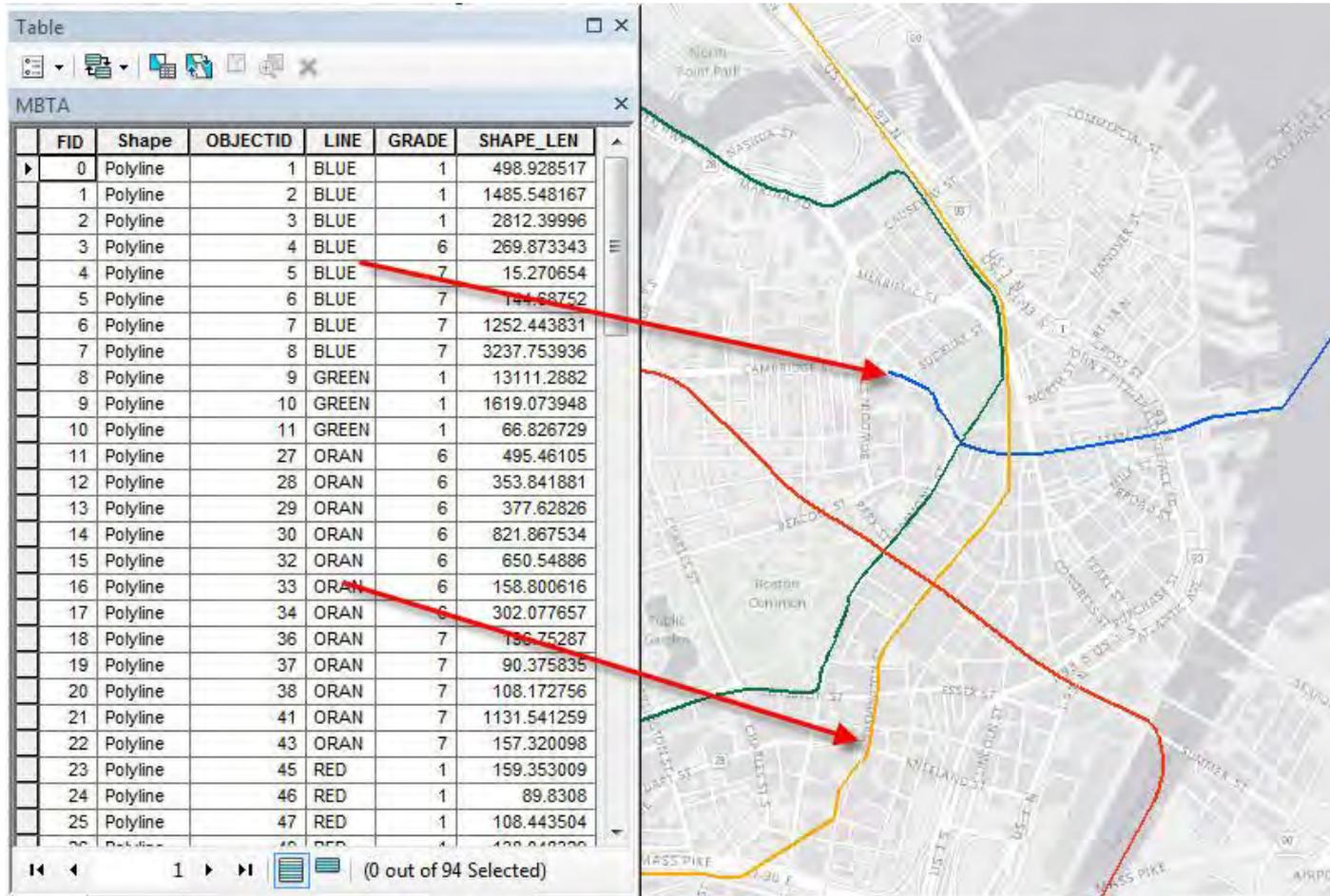
Shapefile

- A vector data storage format developed by ESRI.

- Shapefiles consist of several different files, with extensions such as .shp, .dbf, .prj, etc.

- They are usually downloaded as a .zip file and then need to be unzipped to be used in GIS software.

- Keep ALL the files together in order for them to work in GIS software.

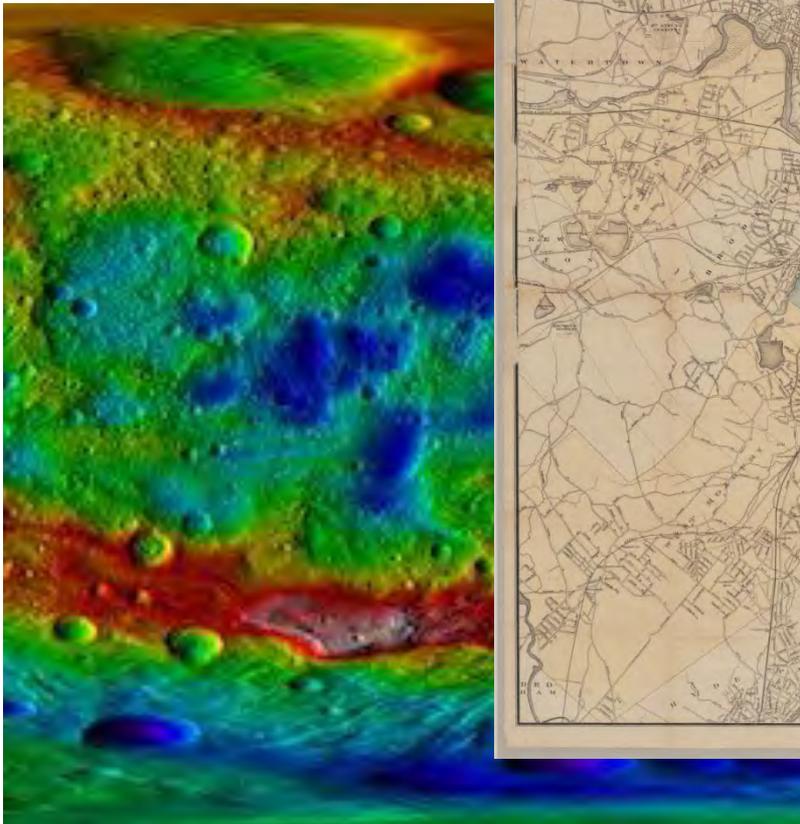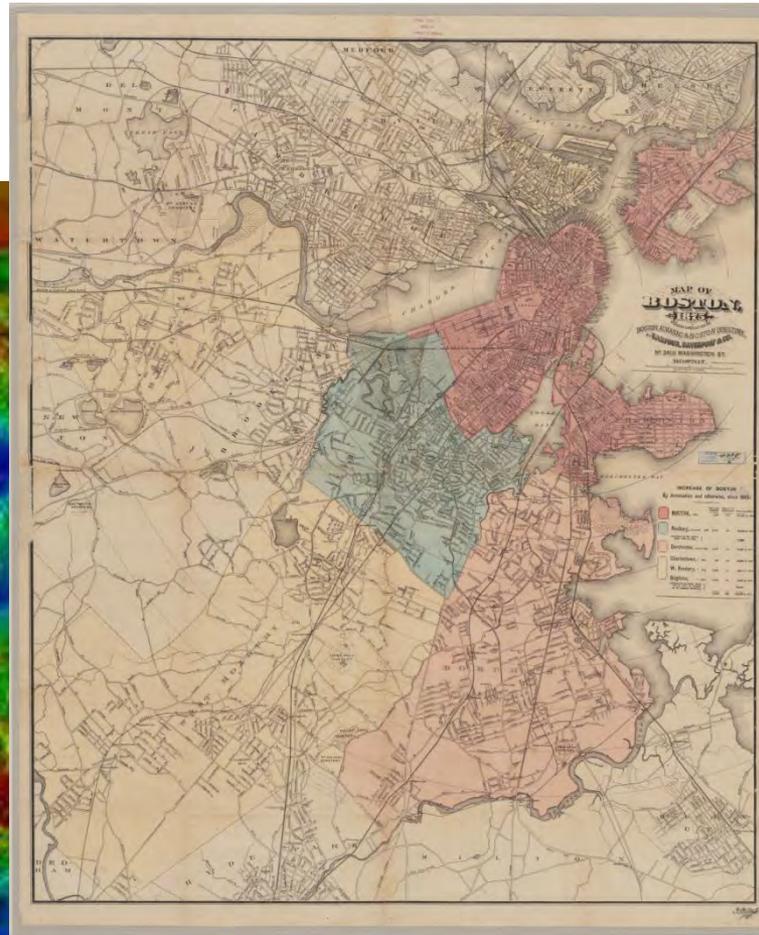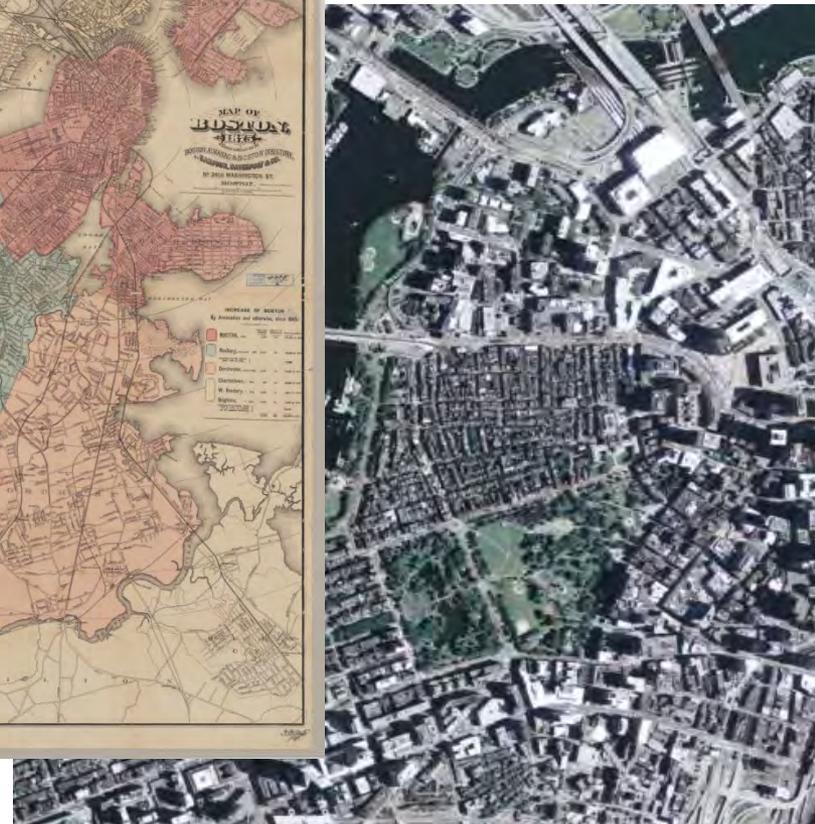SDE_DATA_IN_JAMNAGAR_F7WARD_2011.zip
SDE_DATA_IN_JAMNAGAR_F7WARD_2011.cst
SDE_DATA_IN_JAMNAGAR_F7WARD_2011.dbf
SDE_DATA_IN_JAMNAGAR_F7WARD_2011.prj
SDE_DATA_IN_JAMNAGAR_F7WARD_2011.shp
SDE_DATA_IN_JAMNAGAR_F7WARD_2011.shx

Raster: Pixels

Map reproduction courtesy of the Norman B. Leventhal Map Center at the Boston Public Library.

Image in the Public Domain courtesy of NASA:
http://www.nasa.gov/sites/default/files/pia17037.jpg

11

Jennie Murack, MIT Libraries, January 2016

Tabular

Address	City	State	ZIP Code
1134 Massachusetts Ave	Cambridge	MA	02138
290 Main St	Cambridge	MA	02142
47 Mount Auburn St	Cambridge	MA	02138
428 Massachusetts Ave	Cambridge	MA	02139
314 3rd St	Cambridge	MA	02142
675 W Kendall St	Cambridge	MA	02142
746 Massachusetts Ave	Cambridge	MA	02139
247 Cambridge St	Cambridge	MA	02141
276 Broadway	Cambridge	MA	02139
2370 Massachusetts Ave	Cambridge	MA	02140
1687 Massachusetts Ave	Cambridge	MA	02138
1722 Massachusetts Ave	Cambridge	MA	02138

Geo_FIPS	Total Population	Male	Female
1001020100	1900	944	956
1001020200	2342	1157	1185
1001020300	3297	1451	1846
1001020400	4272	2056	2216
1001020500	10881	5202	5679
1001020600	3782	1769	2013
1001020700	2799	1140	1659
1001020801	3096	1475	1621
1001020802	10471	5283	5188
1001020900	5637	2908	2729

Latitude	Longitude
42.370882	-71.114246
42.362318	-71.086001
42.371406	-71.116643
42.363327	-71.100944
42.363039	-71.08254
42.3364681	-71.082363
42.366312	-71.105351
42.371092	-71.080076
42.367804	-71.097311

Software for Viewing Spatial Data

- GIS Software (ArcGIS, QGIS)

- Online Mapping Software (ArcGIS Online, CartoDB, Mapbox)

- Statistical/Data Visualization Software (Tableau, R, MatLab)

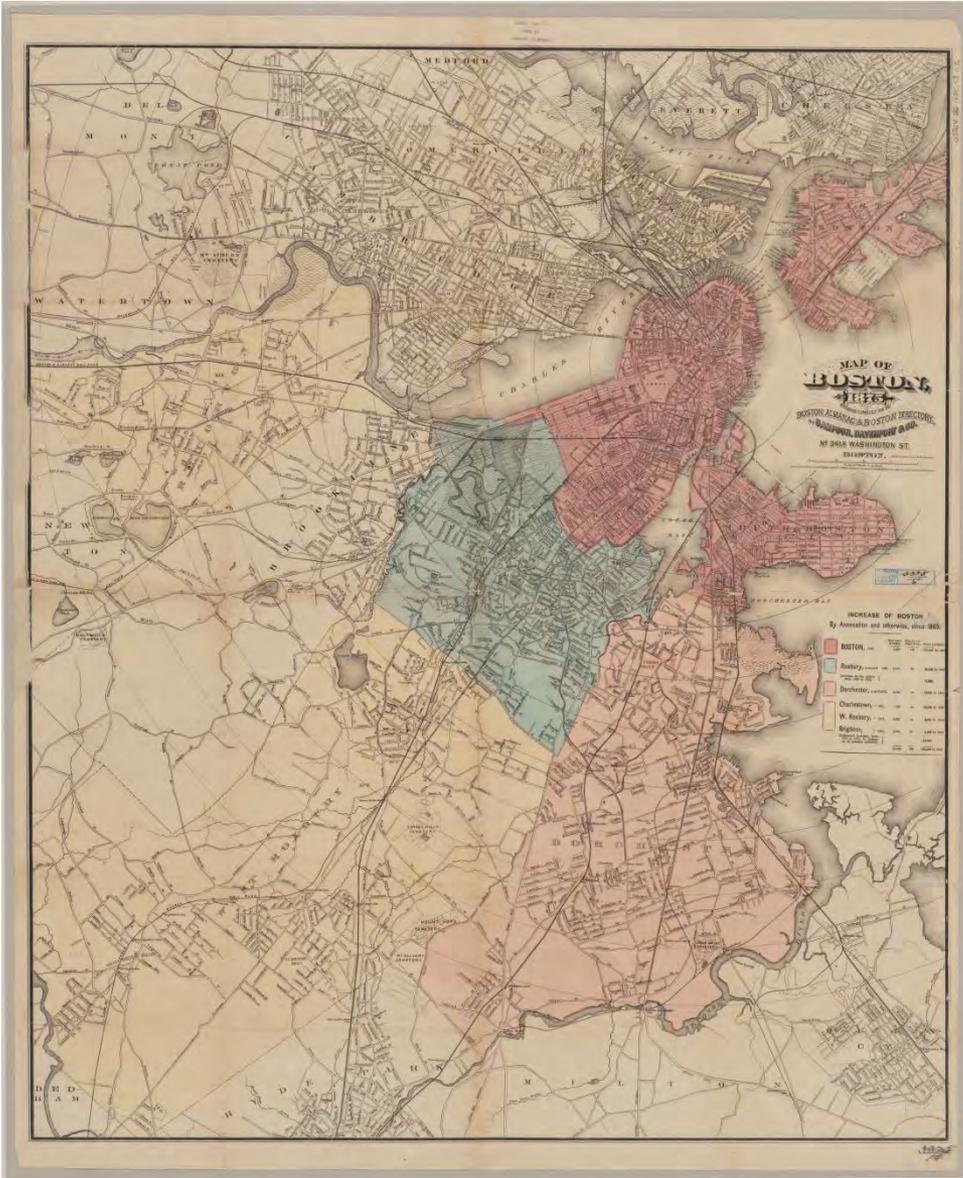- Programming Languages (Python, Javascript (D3, leaflet), R, MatLab)

Is this spatial data?
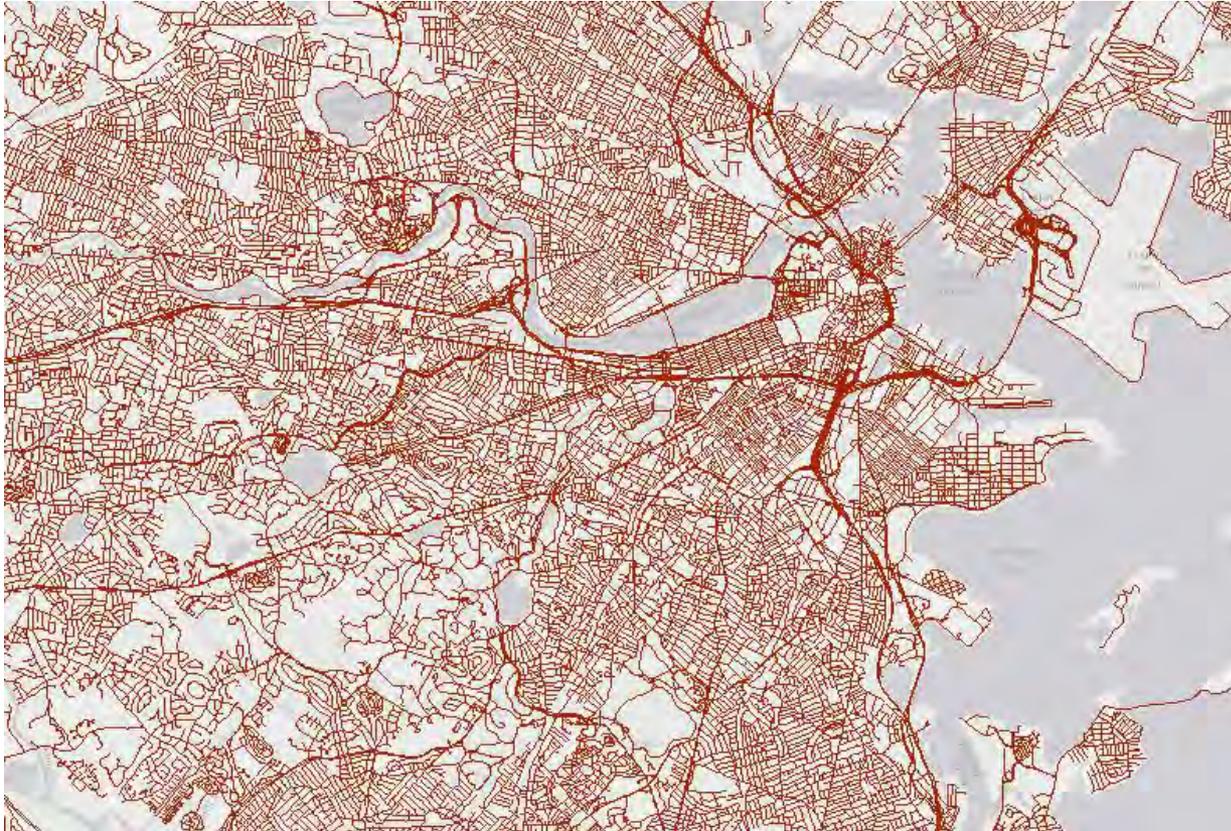
Yes, these are lines representing roads in the Boston area (vector data)

Is this spatial data?

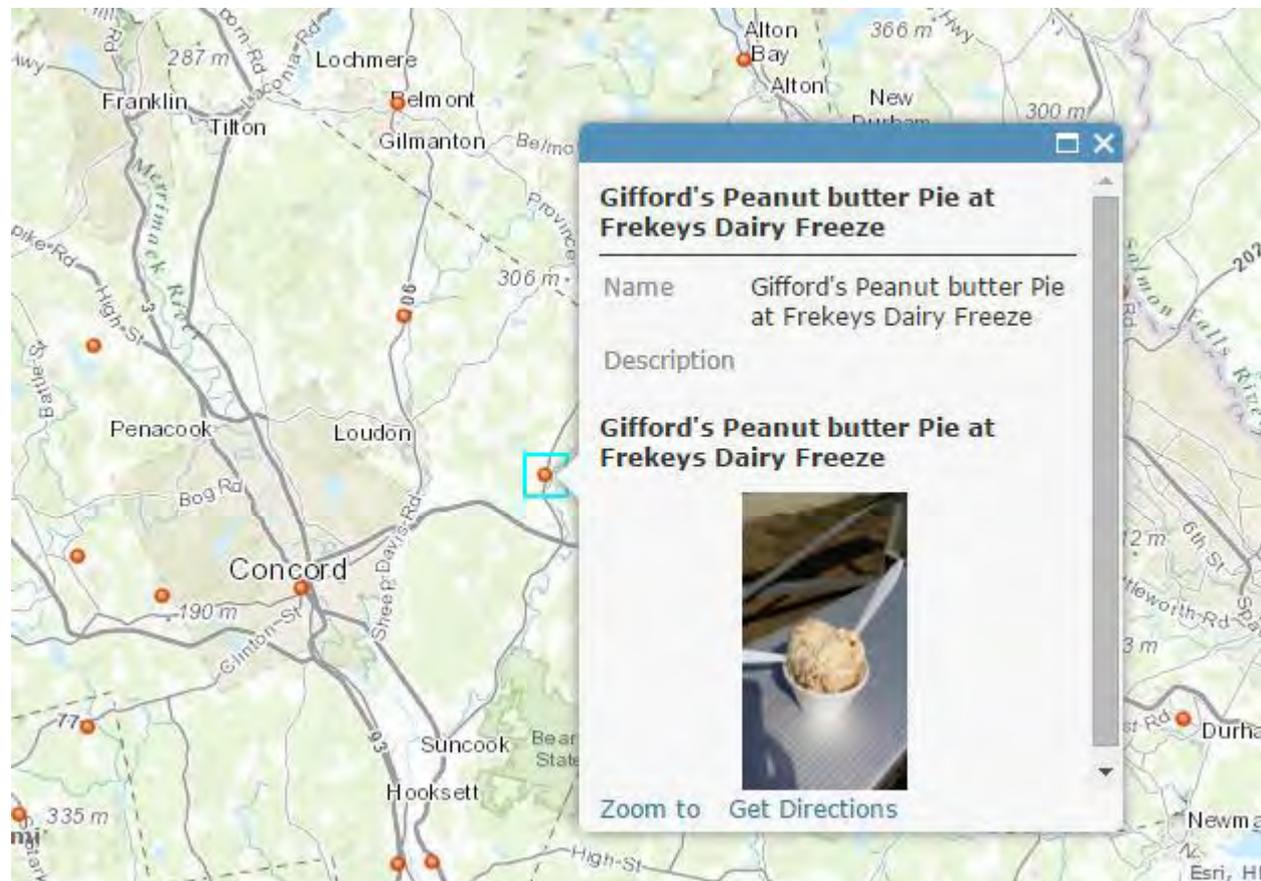Yes, this is a scanned, historic map of the Boston area (raster data)

Map reproduction courtesy of the Norman B. Leventhal Map Center at the Boston Public Library.

Jennie Murack, MIT Libraries, January 2016

15

Is this spatial data?

It could be. See the next slide.

Is this spatial data?

Yes. When an image is georeferenced to a map or has embedded coordinate data, it is spatial data. (raster data)

Is this spatial data?

No, there is no information you could use to map this data. If the location of each business were listed, it would be tabular spatial data.

Location Sales Volume Actual	Type of Business
$104,000	Private
$52,000	Private
$157,000	Private
$157,000	Private
$104,000	Private
$104,000	Private
$157,000	Private
$104,000	Private
$52,000	Private
$104,000	Private
$104,000	Private
$104,000	Private
$157,000	Private
$209,000	Private

Is this spatial data?

Yes. It doesn't use the same coordinate system as data that is plotted on the earth, but you could map items on this brain. This scan is raster data since it uses pixels.
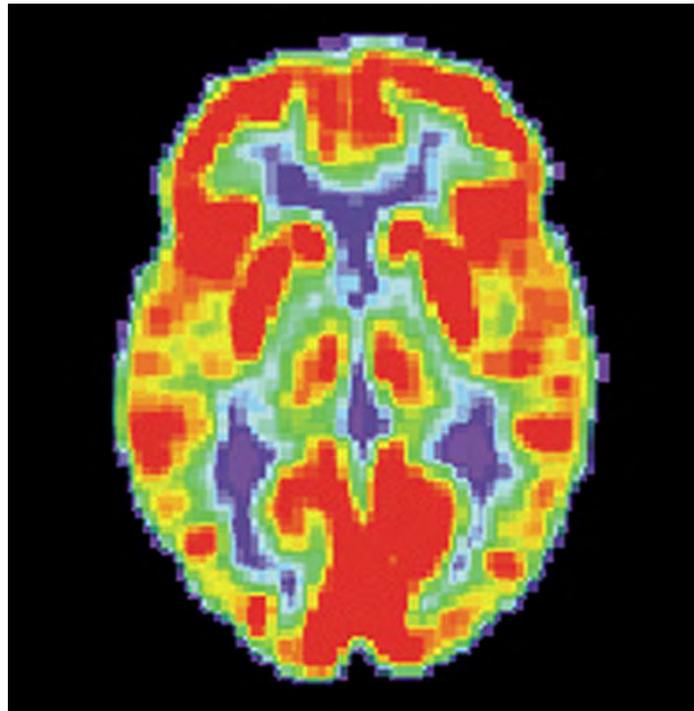
This Image in the Public Domain Courtesy of NIH https://en.wikipedia.org/wiki/Brain _positron_emission_tomography#/media/File:PET_Normal_brain.jpg.

CHAPTER II.

The world was all before them, where to choose
Their place of rest, and Providence their guide.

Milton.

At the time when that resolution was taken, very little was known of the state of this part of America. Columbus, who was a Genoese, had discovered the West Indies in 1492, under Spanish patronage. In the spring of 1497. John Cabot, a Venetian, obtained a commission from Henry VII. under which he made a voyage from England, and discovered the continent of North America, in the latitude of forty-five degrees. He proceeded first northerly to the sixty-seventh degree, and then returning coasted to the south as far as Florida. The southern continent was found by Columbus in 1498. The first voyage from France to America was made in 1524, by Verrazano, a Florentine. By these voyages and discoveries, the princes of the several states, which authorized them, fancied themselves to have acquired the rights of territory and jurisdiction. The Spanish and French made various settlements in the course of a century. On the part of the English little or nothing seems to have been done, until Sir Walter Raleigh attempted the colonization of Virginia, between 1584 and 1607, when Jamestown was established. An effort was made to plant a colony near the mouth of the river Kennebeck, under the auspices of some very able persons, about the year 1607, but the severity of the winter and the loss of their storehouse by fire disheartened them, and the men took advantage of the first shipping to return to England, the next year. A settlement was commenced at Newfoundland in 1610, and in 1614 some Dutch adventurers built a fort at Albany, and laid the foundation of that city, and of New-York.

Massachusetts Bay, the bottom of which forms the harbour of Boston, had not been much explored previous to that time. Twelve years before, Bartholomew Gosnold, the first Englishman who came in a direct course to this part of America,

Is this spatial data?

No, not in its current form. However, you could analyze this text and put the spatial locations in table to map them, or plot them manually on a map to transform this text into spatial data.

GET YOUR DATA

Data Sources

- MIT sources
  - GeoWeb: use any web browser (includes data downloads as well as DVDs and Maps in the library)
- Internet
  - http://libguides.mit.edu/gis (Links to data sources)
  - Google search
  - OpenStreetMap.org
- Create your own
  - GPS, digitizing, etc.

Not finding what you want? GIS data purchase requests? Contact GIS Help.

GeoWeb – search 2000+ layers of MIT hosted GIS data and data from other schools.

GIS Services links to data

OpenStreetMap.org

Open data! Anyone can contribute and download.

Jennie Murack, MIT Libraries, January 2016

OpenStreeMap

- Data cannot be downloaded in a shapefile format from OpenStreetMap, but you can do this from other websites.

- Use the Trimble Data Marketplace: http://data.trimble.com/market/index.html

- See this video for instructions: https://youtu.be/K8nUs-Hf4Fc

Search Tips

- Use "GIS", "data", or "map" as some of your search terms.
  - "wind data gis", "US census gis"
- If you are looking for data from a specific location, search for a town/county/state/country GIS organization and contact them.
  - "Cambridge MA GIS department"
- Many cities have open data portals that contain GIS and other data.
  - "Boston open data"
- When searching for foreign data, search in the language native to that country.
- Look for universities near your area of interest and contact researchers. They may be willing to share their data!

ACTIVITY:
- FIND A SPATIAL DATA FILE USING ANY SOURCE.
-KEEP THE SOURCE/WEBSITE FOR YOUR DATA OPEN.

ACTIVITY: OPEN THE DATA IN CARTODB

Connecting to a dataset:

http://docs.cartodb.com/cartodb-editor/datasets/#connect-dataset

Geocoding tabular data:

http://docs.cartodb.com/tutorials/how_to_georeference/

ACTIVITY: OPEN THE DATA IN QGIS

Add data layers:

https://docs.qgis.org/2.2/en/docs/user_manual/working_with_vector/supported_data.html

Add tabular data with XY coordinates:

http://www.qgistutorials.com/en/docs/importing_spreadsheets_csv.html

Geocode addresses:

http://www.gislounge.com/how-to-geocode-addresses-using-qgis/

DID YOU NOTICE ANYTHING UNEXPECTED ABOUT YOUR DATA?

EVALUATING SPATIAL DATA

Metadata

- Information about the data layer

- Read the metadata to determine who created the data, when it was created, what the codes in the table mean, if there are constraints on how it can be used, etc.

- You can find metadata:
  - Downloaded with your data layers
  - On the website where you got your data
  - Sometimes you may need to contact the data provider to get metadata

- Metadata is most commonly in html/xml format, text files, or in a table format, such as excel or csv.

Metadata Standards

- FGDC (Federal Geographic Data Committee)

- ISO (International Organization for Standardization)

- List how the data should be described.

- Include standard headings/tags for each element about the data.

1.1 Citation -- information to be used to reference the data set.
    Type: compound

    Short Name: citation

1.2 Description -- a characterization of the data set, including its intended use and limitations.
    Type: compound

    Short Name: descript

    1.2.1 Abstract -- a brief narrative summary of the data set.
        Type: text

        Domain: free text

        Short Name: abstract

    1.2.2 Purpose -- a summary of the intentions with which the data set was developed.
        Type: text

        Domain: free text

        Short Name: purpose

    1.2.3 Supplemental Information -- other descriptive information about the data set.
        Type: text

        Domain: free text

        Short Name: supplinf

Metadata

Jennie Murack, MIT Libraries, January 2016

Metadata

Entity and Attribute Information

SDE_DATA.INT_A1CNTRY_2005

ObjectID

Attribute Definition	Internal feature number.
Attribute Definition Source	ESRI

FIPS_CNTRY

Attribute Definition	The FIPS code (two-letter) for the country.
Attribute Definition Source	Department of Commerce, National Institute of Standards and Technology

GMI_CNTRY

Attribute Definition	The country code (three-letter) for the country from Global Mapping International.
Attribute Definition Source	Global Mapping International

ISO_2DIGIT

Attribute Definition	The country code (two-letter) for the country from the International Organization for Standardization.
Attribute Definition Source	International Organization for Standardization

ISO_3DIGIT

Data Quality

- Lineage
- Accuracy (spatial, temporal, thematic/attribute)
- Consistency
- Completeness

Lineage

Street Centerlines

Feature Layer by BostonGIS
Source: Feature Service
Last Modified: May 19, 2015

(0 ratings, 41 views)

Sign in to rate this item.

Facebook   Twitter

OPEN ▾

Description

Layers

Street_Centerlines ▾

Properties

Tags	boston, bostonmaps, streets, centerlines, test, tpl
Credits	
Size	31 MB
Extent	Left: -71.19 Right: -70.92
	Top: 42.4 Bottom: 42.23

Where did the data come from? How as it derived?

MassGIS Anadromous Fish

Overview

The Department of Fish and Game (DFG) GIS Program working in conjunction with biologists from the MA Division of Marine Fisheries and the MA Division of Fisheries and Wildlife compiled and automated a point coverage of anadromous fish data. The data include all known coastal anadromous fish runs spawning habitat and runs for three

Methodology

During interviews with Division biologists the data points were compiled onto 1:25000 basemaps using the best available hydrographic data. "Heads-up" digitizing was used to automate the data points. Paper forms were completed with information regarding locations of barriers, fishways, beginning and ending of runs and spawning habitat. The information from these forms was used to populate the associated look up tables (RIV, RUN, BAR).

Accuracy

- Is the data in proper place and time? Are the attributes accurate?

- If the data appear to be the correct shape, but are showing up in an unexpected location on your map, check the map projection.
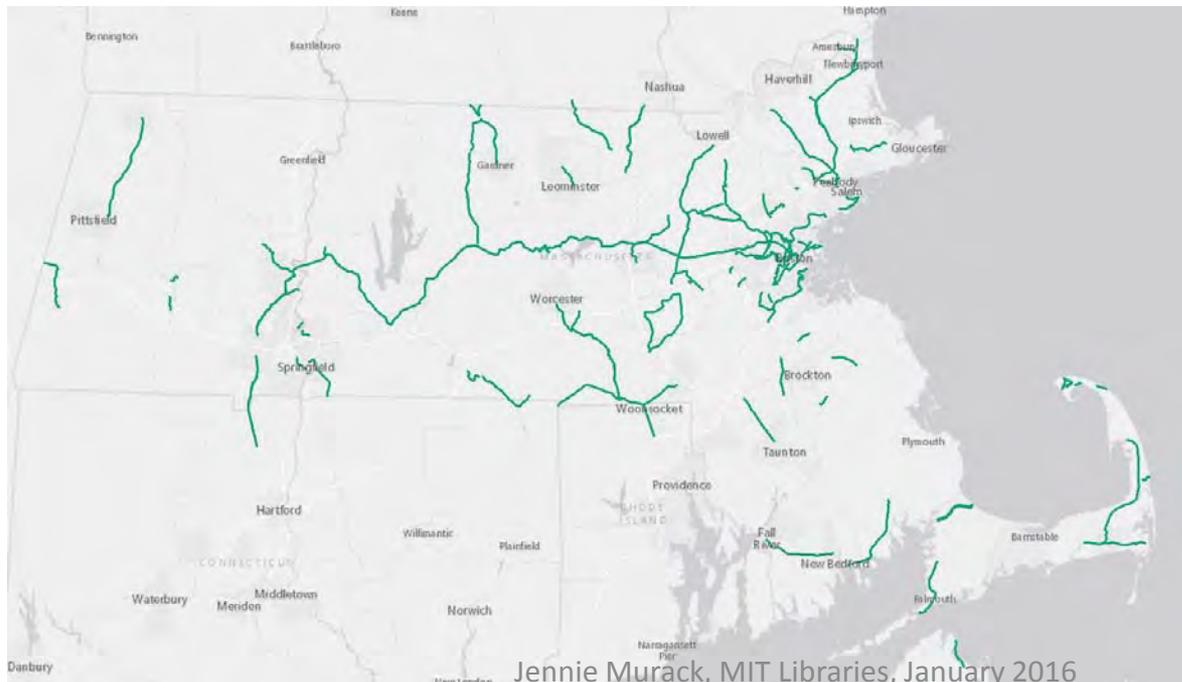
Consistency

- The absence of contradictions.
- Data follow topological rules (polygons to do overlap, lines intersect at nodes, etc.)
- Appropriate attributes.

| ID	State
1	Alabama
2	Connecticut
3	Massachusetts
4	California
5	Middlesex

Completeness

- Does the data include all relevant features and attributes?

- Consider the time period.

If this data layer was called "New England bicycle trails" it would not be complete. It was last updated in 2002 and is complete for that time period in Massachusetts.

Jennie Murack, MIT Libraries, January 2016

41

Is this the right data for your project?

- Consider scale, type of features, how data are represented, etc.
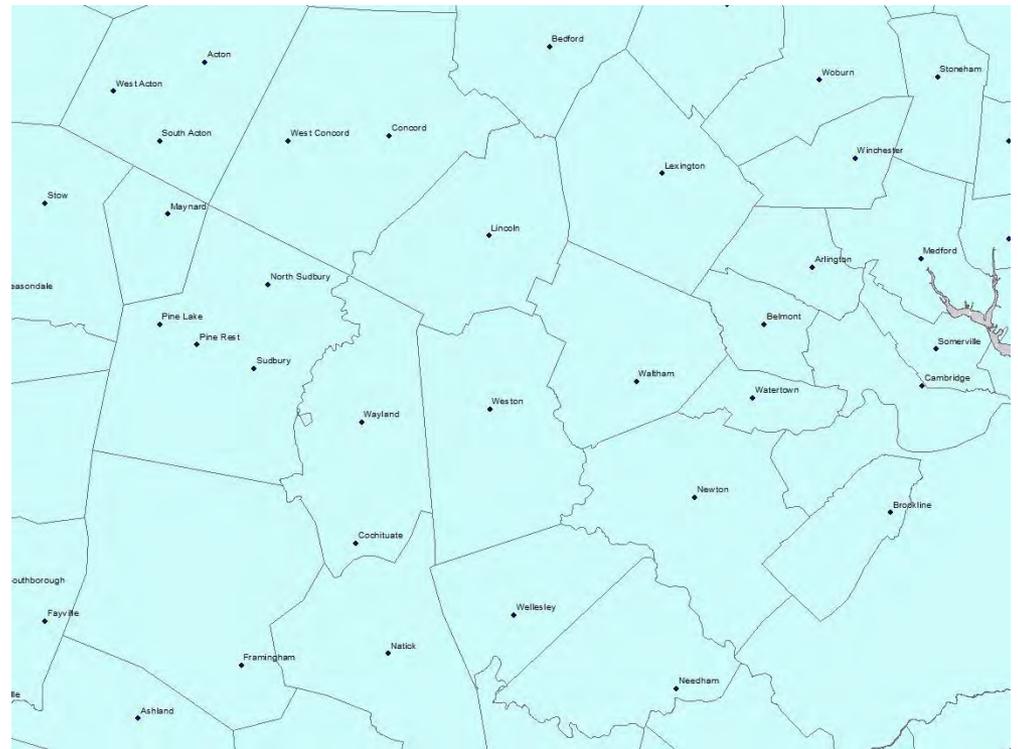
ACTIVITY: FIND THE METADATA FOR YOUR DATA

DATA MANAGEMENT TIPS

Data Management Tips

GIS projects tend to generate **many files**, which are generally **large in size**. For file naming:

- Use file names that represent the file (default names like Export_Output are not helpful if you need to come back to your project later).

- Some software programs and tools may have file name constraints (e.g. an eight character limit without spaces). Watch out for this with ESRI ArcToolbox.

- Backup Your Data!

Data Management Tips

Keep detailed notes about:

- Data sources
- Licensing constraints
- Data processing steps
- What is stored where
  - The GIS project maintains links to the individual data files (the data is not embedded in the map document itself)
  - GIS formats, like shapefile (SHP), have many files that are linked together and must stay together in order to function
- Descriptions of the files you create and use (ArcCatalog has built-in tools for creating and editing metadata)

RES.STR-001 Geographic Information System (GIS)Tutorial
January IAP 2016