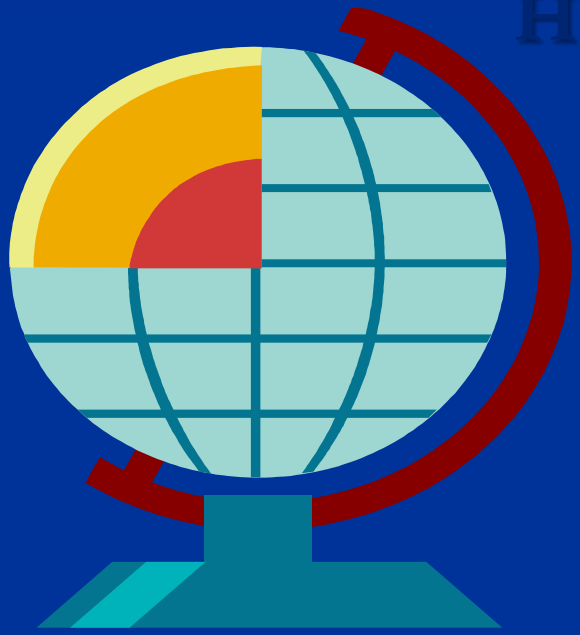
Geospatial Data Concepts

Center for Geographic Analysis

Harvard University



Nicole Alexander, Ph. D.

nalexander[at]cga.harvard.edu

Geospatial Data Concepts

- How Geospatial Data are Structured
- Where Geospatial Data Come From
- Key Issues in Managing Geospatial Data

Geospatial Data

Points

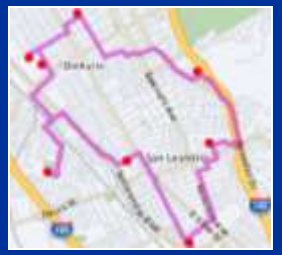
Lines

Polygons

3D

- Represent locations, places or spaces on Earth – geographic information
- Can be displayed on a map

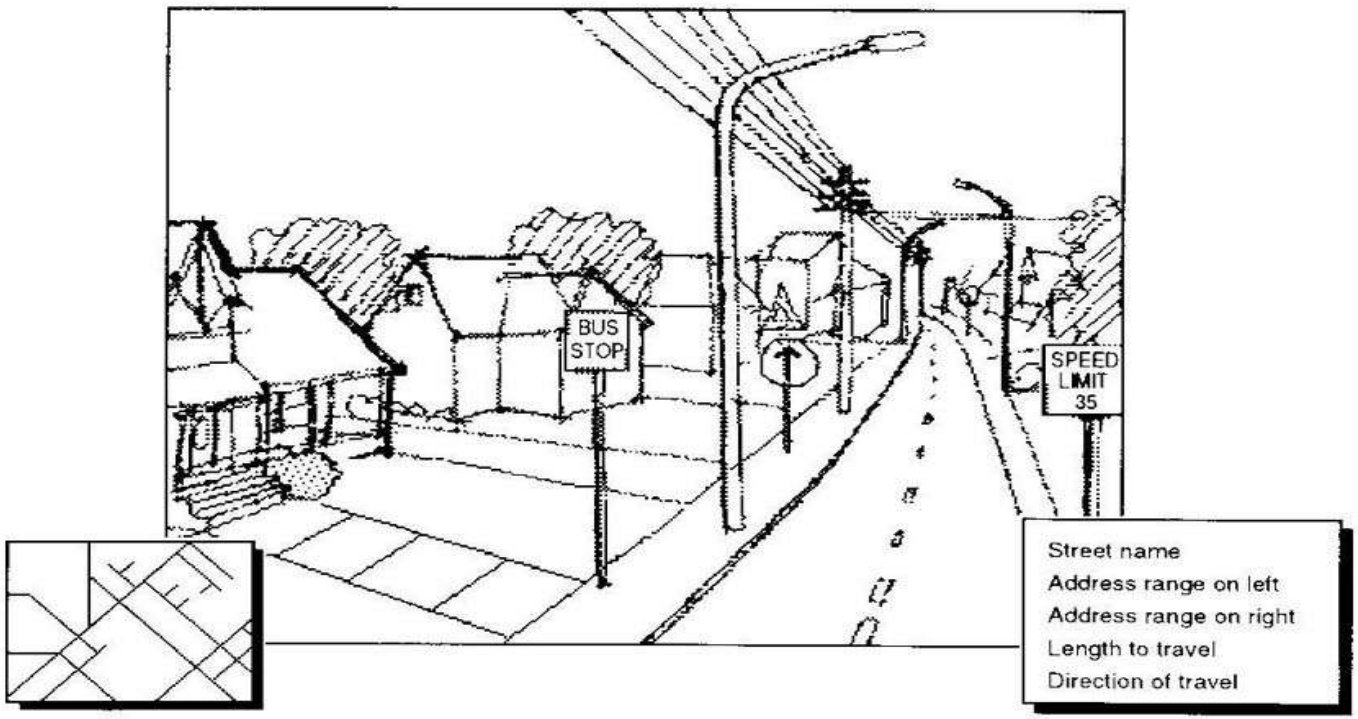
Satellite Imagery



Networks

Harvard University
Cambridge, MA 02138
Addresses

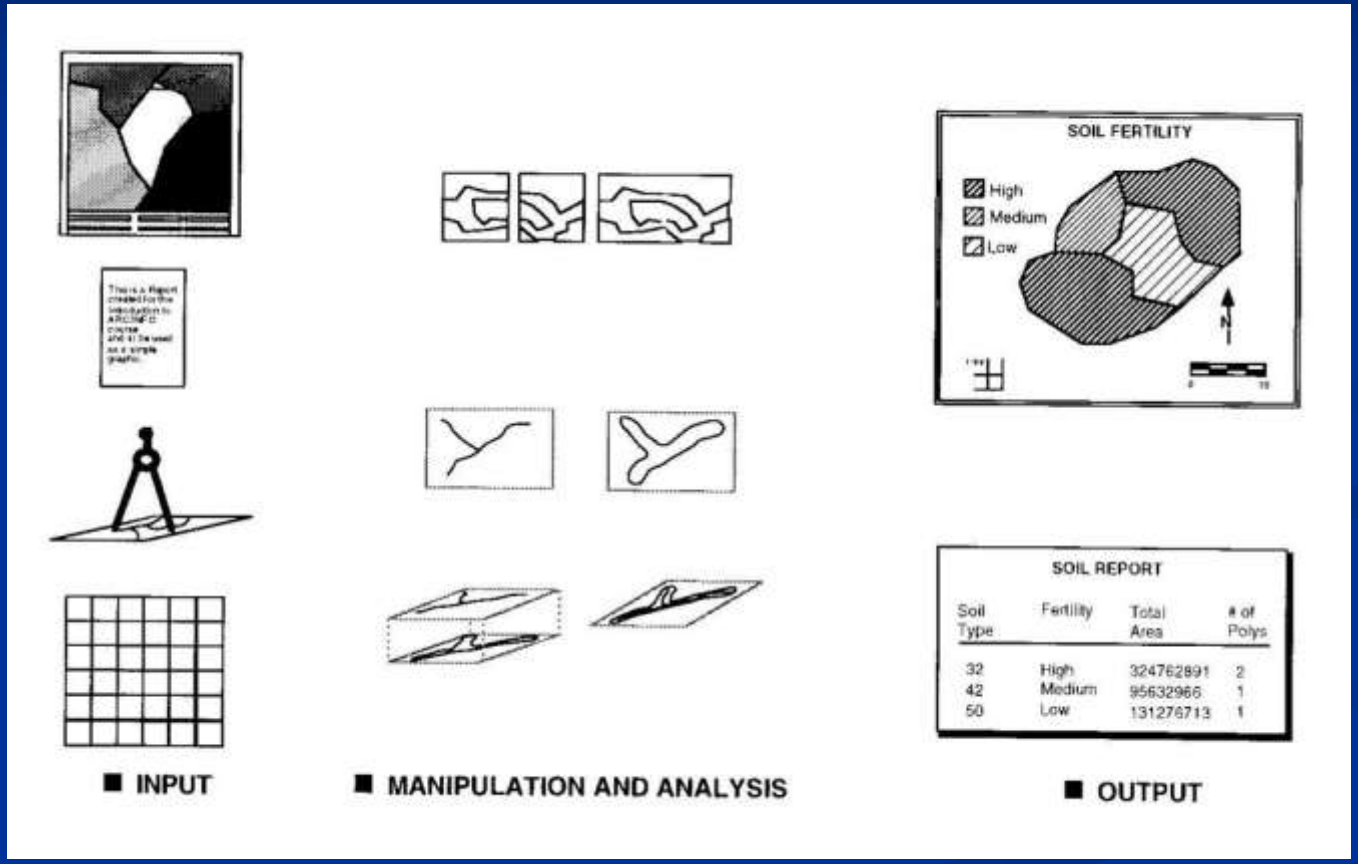
Geospatial Data



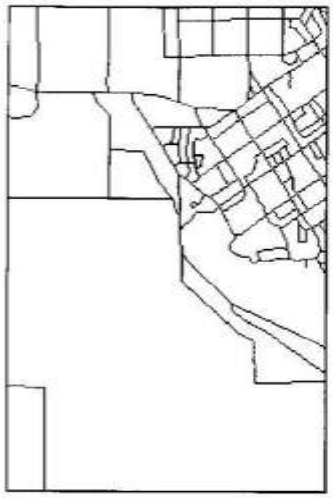
Geographic Information System (GIS) Definition

- A computer system for capturing, storing, checking, manipulating, analyzing, and displaying data which is spatially referenced to the Earth (DoE 1987; Maguire, Goodchild and Rhind 1991)

What you can do with GIS:



Display Data

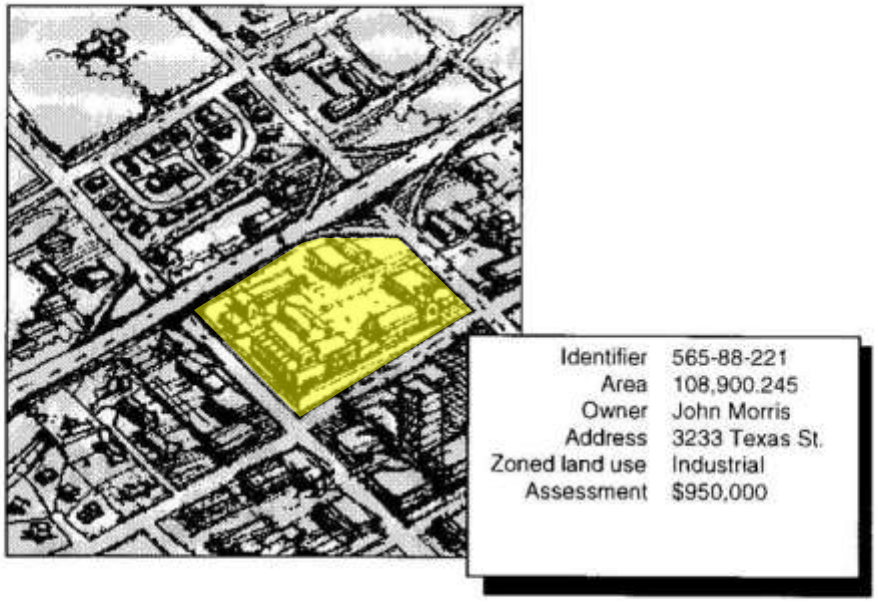


■ SPATIAL DATA

```
1
FNODE#          =      1
TNODE#          =      2
LPOLY#          =      1
RPOLY#          =      3
LENGTH          =           2652.82976
STREETS#        =      1
STREETS-ID      =      0
SLINE_FLAG      =
STRT_CODE       =  0
COUNTY          =  0
CITY            =  0
ADDRESS         =
WALK_IMPED      =      0.000
FT_DRV_IMP      =      0.000
TF_DRV_IMP      =      0.000
DIRECTION       =  0
2
FNODE#          =      2
TNODE#          =      3
LPOLY#          =      1
RPOLY#          =      4
LENGTH          =           2056.20654
STREETS#        =      2
STREETS-ID      =      0
SLINE_FLAG      =
STRT_CODE       =  0
COUNTY          =  0
CITY            =  0
ADDRESS         =
WALK_IMPED      =      0.000
FT_DRV_IMP      =      0.000
TF_DRV_IMP      =      0.000
DIRECTION       =  0
```

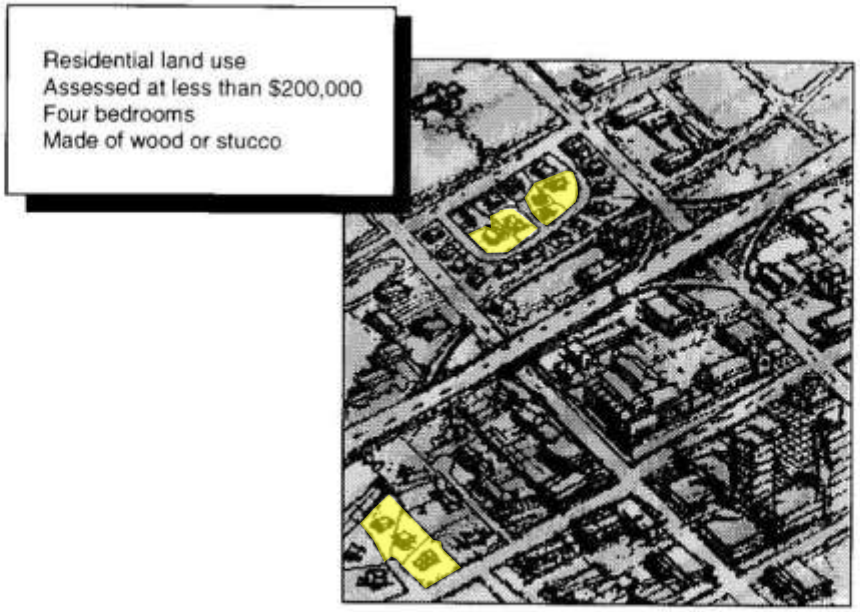
■ ATTRIBUTE DATA

Locate and Identify Features



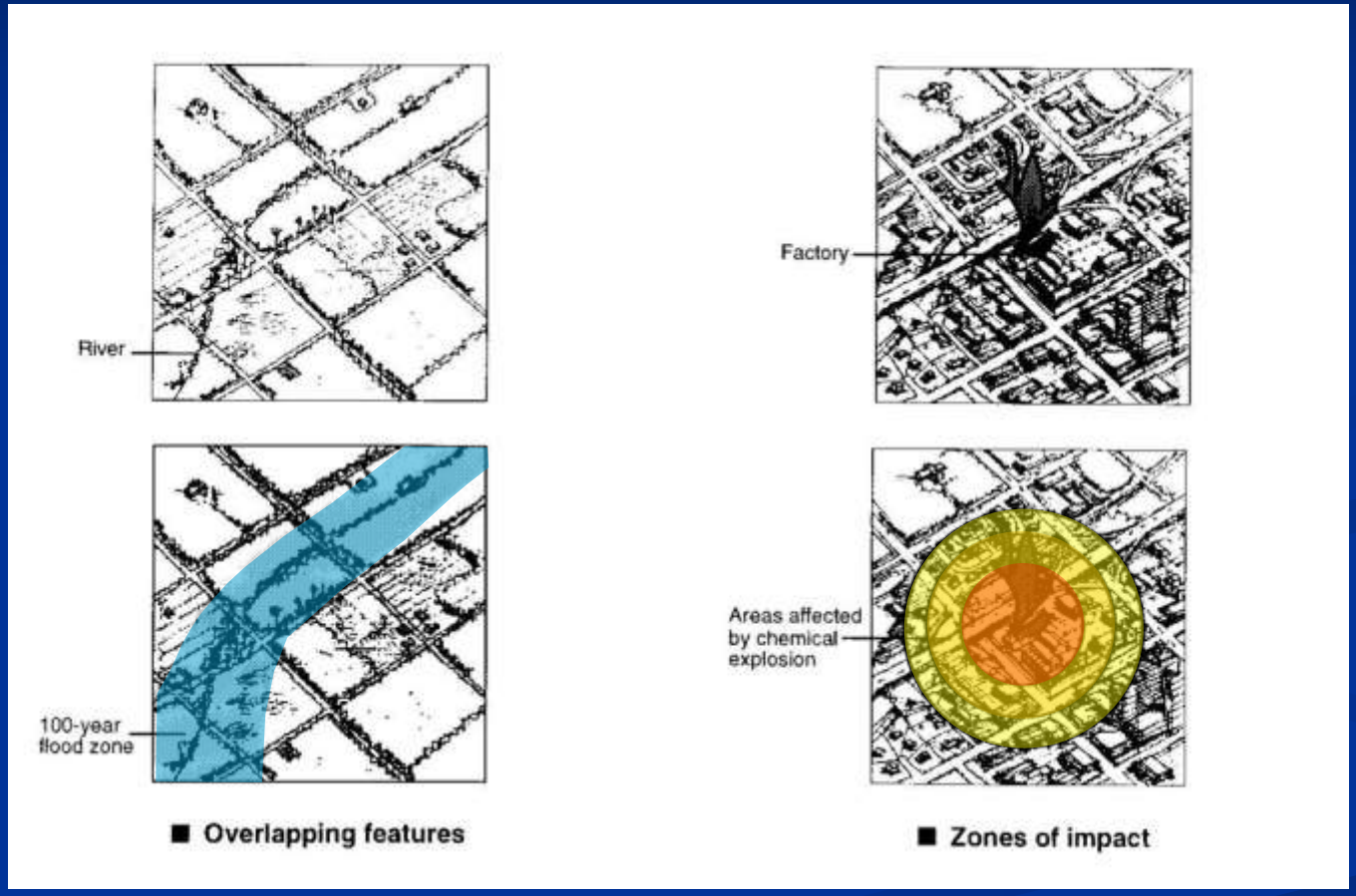
- Who owns the lot at 3233 Texas Street, and what is the lot's zoning?

Specify Conditions



- Where are houses located that you might consider buying?

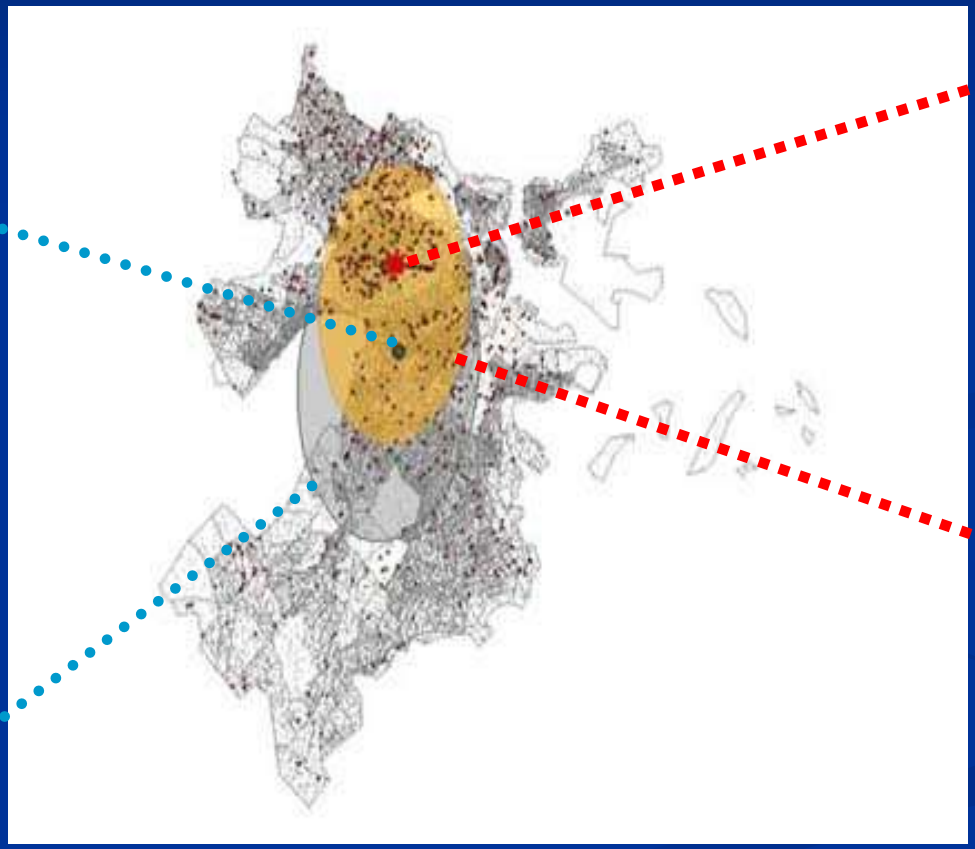
Perform Geographic Analysis



Reveal Patterns and Processes in Geographic Information

Mean Population Weighted Center

(close to intersection of Fenway and Forsyth in Boston)



Hazard Area and Oil Spills Center

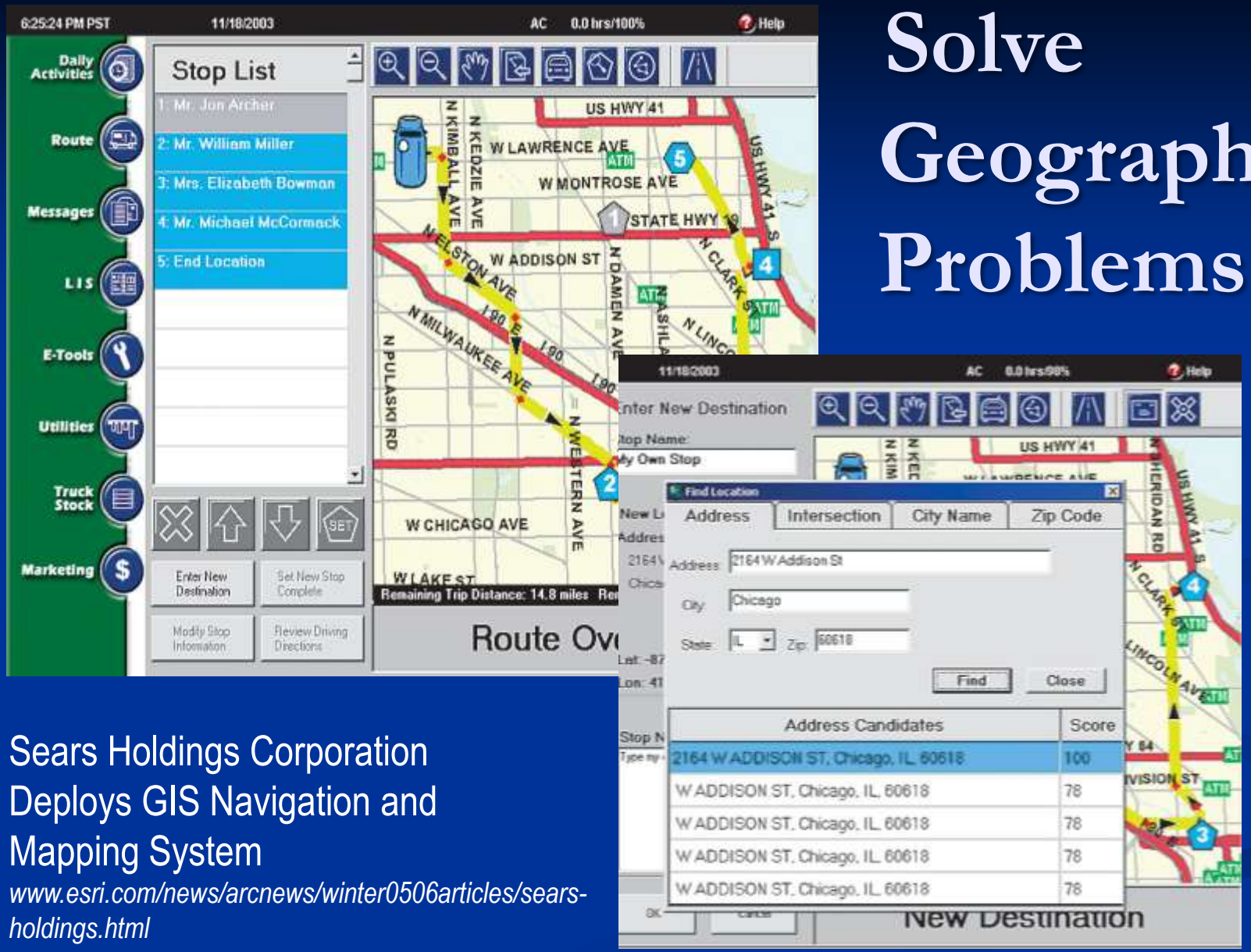
(close to intersection of Main and Portland in Cambridge)

Directional Ellipse for Spills

Directional Ellipse for Population

Oil Spills in Boston, Cambridge, Somerville

Solve Geographic Problems



Sears Holdings Corporation Deploys GIS Navigation and Mapping System

www.esri.com/news/arcnews/winter0506articles/sears-holdings.html

Geographic Information Science

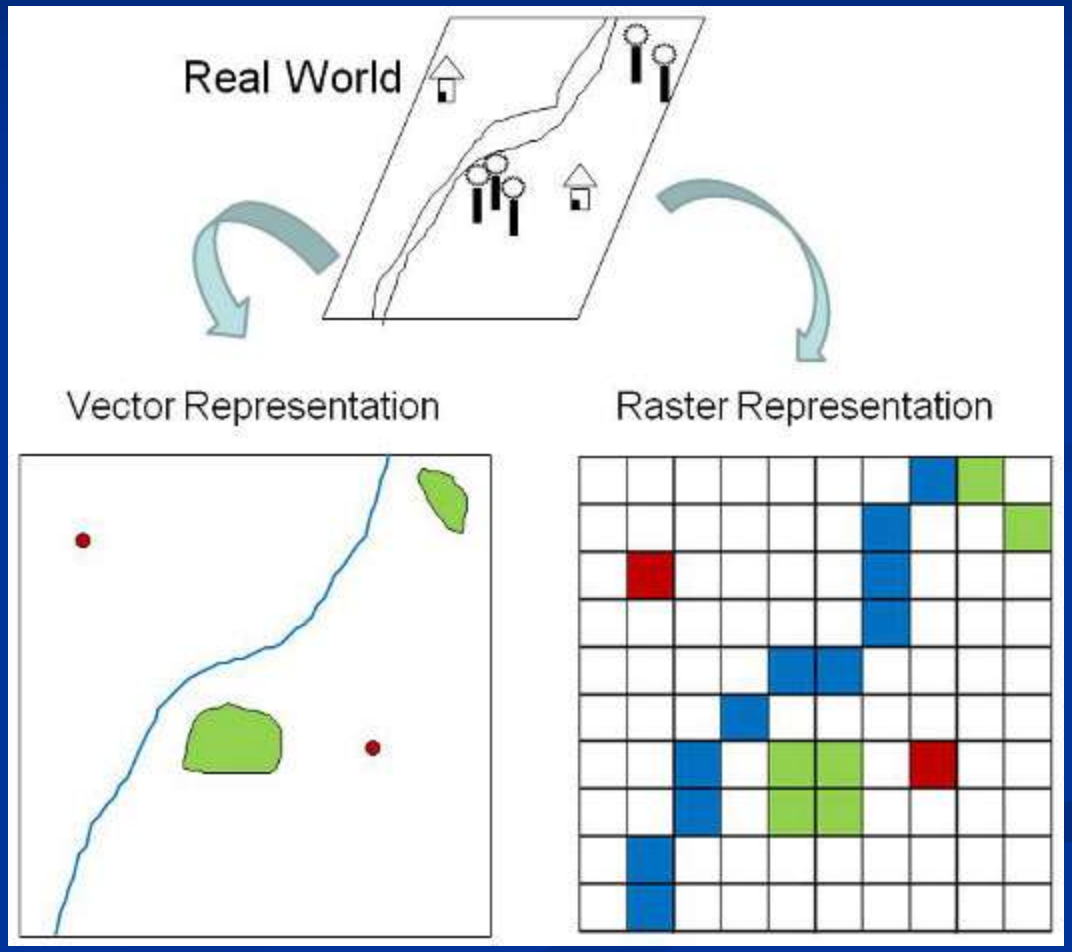
- The academic theory behind the development, use and application of GIS

GIS Data Models

- A data model is a mechanism for describing reality
- GIS data modeling is the process of representing specific aspects of the real world in a GIS
- *Raster* and *Vector* are two common GIS data models

GIS Data Models

- Choose a data model that closely represents your phenomenon of interest



Vector Data Model

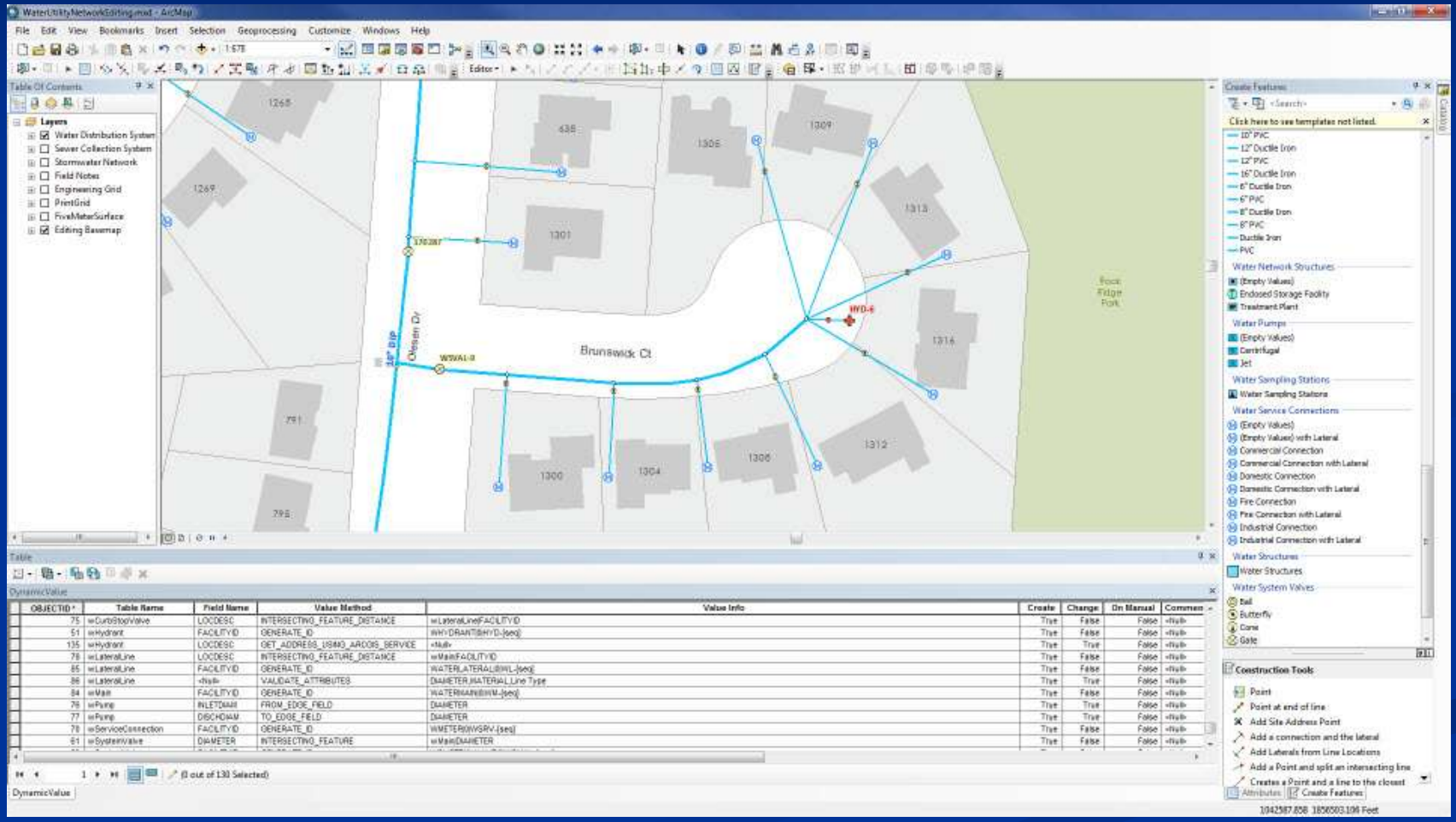
- Used when individual coordinates are important
- Captures discrete features with precise shapes and boundaries
- Provides a more concise spatial description
- Common vector formats: Shapefile, Geodatabase, KML and GML

Vector Attribute Data

- Stored in tables
- Each row represents a spatial feature and each column describes a characteristic

Record	Soil-ID	Area	Perimeter
1	1	106.39	495.86
2	2	8310.84	508382.38
3	3	554.11	13829.50
4	4	531.83	19000.03
5	5	673.88	23931.47

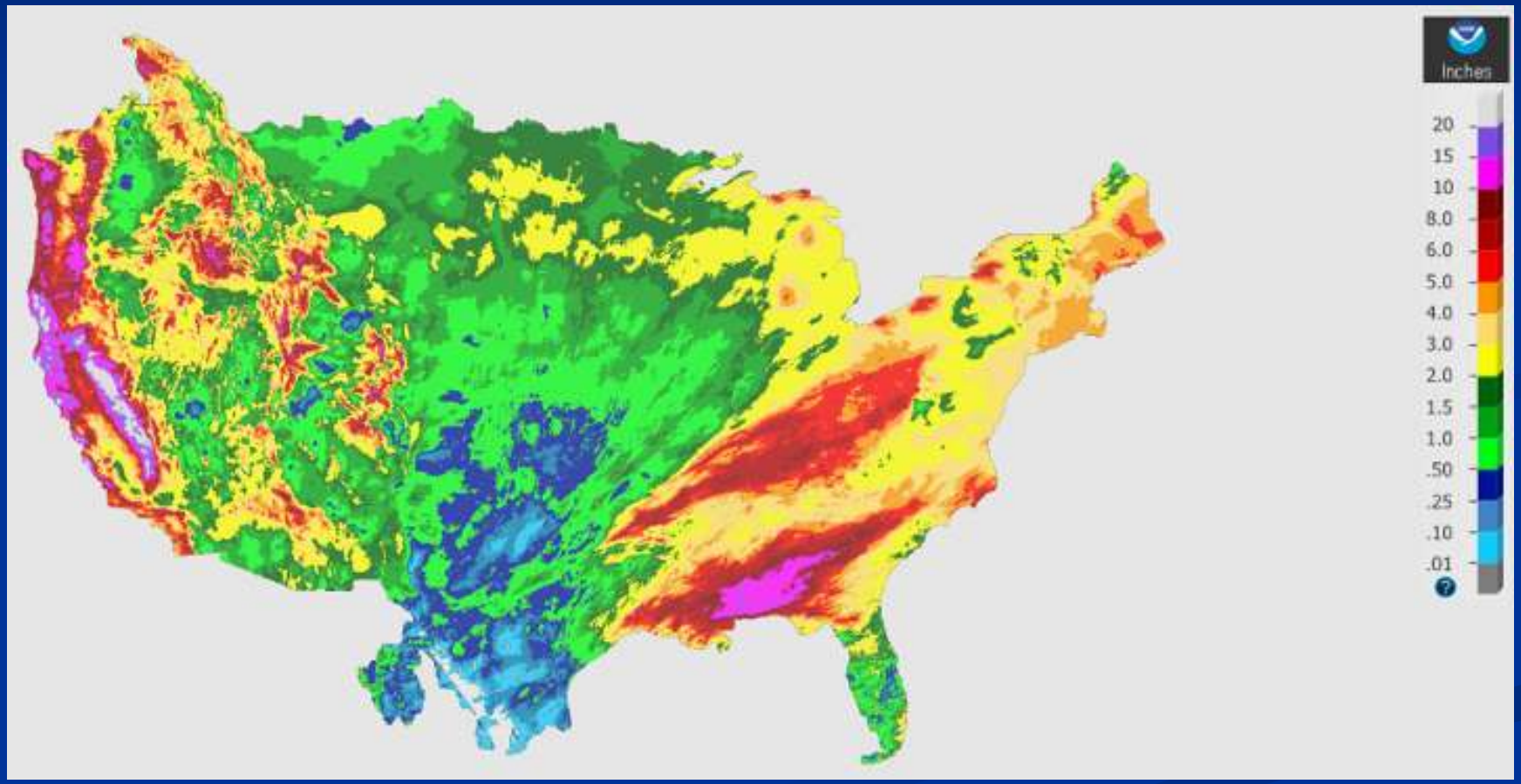
Water Utility Application



Raster Data Model

- Used to model continuous phenomena and images of the earth
- Data becomes more voluminous as cell size decreases
- Data can be compressed
- Easy to display and analyze
- Common raster formats: GeoTIFF, JPEG2000, WorldFile, MrSID, ECW, IMG and Grid

Weather Application



NOAA: January 09, 2017 30-Day Observed Precipitation

NOAA

Where Data Come From

- How data are captured determines the quality of decisions that can be made from analyzing the data
  - *Primary* sources: obtained through direct measurement
  - *Secondary* sources: derived from other sources

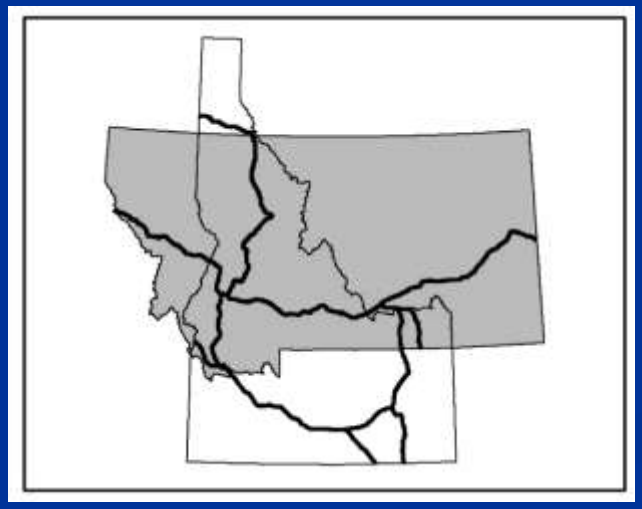
GIS Data Sources

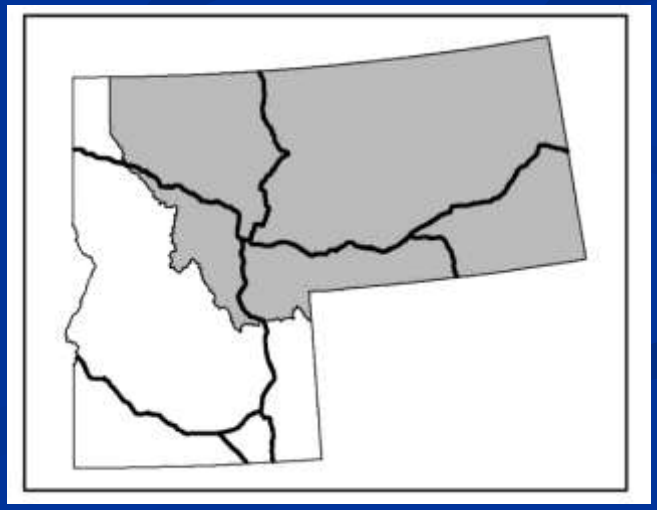
	RASTER	VECTOR
Primary	• Digital satellite remote-sensing images • Digital aerial photographs	• GPS measurements • Field survey measurements • LIDAR
Secondary	• Scanned maps and photographs • Digital elevation models from topographic map contours • Rasterizing vector data	• Photogrammetry • Topographic maps • Toponymy (place-name) databases • Digitizing • COGO • Vectorizing raster data

Coordinate System

- A reference system used to represent the locations of geographic features within a common geographic framework
- Two map layers will not register spatially if they are in different coordinate systems

Layers using Different Coordinate Systems





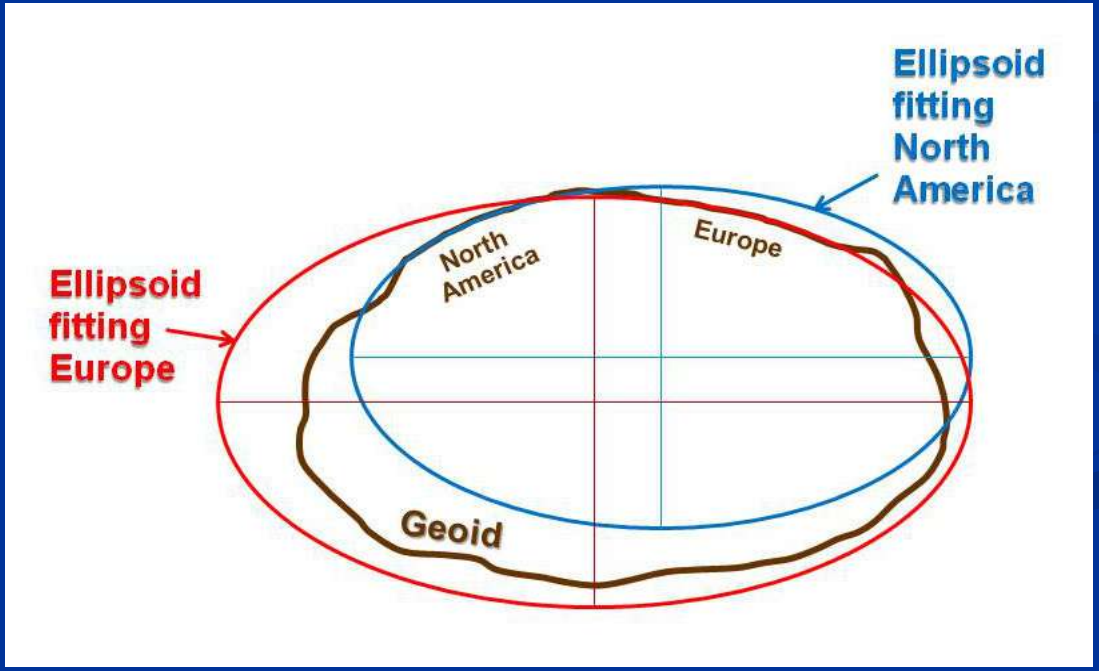
Layers using Same Coordinate System

Types of Coordinate Systems

- Geographic (Geodetic) coordinate systems (GCS)
  - Based on a spherical (3D) model of the Earth
  - Include a *geodetic datum*: a reference ellipsoid, its origin, and orientation
  - latitude, longitude
- Projected coordinate systems
  - Geographic coordinate systems projected unto a plane (or flat surface)
  - GCS + *map projections* and projection parameters

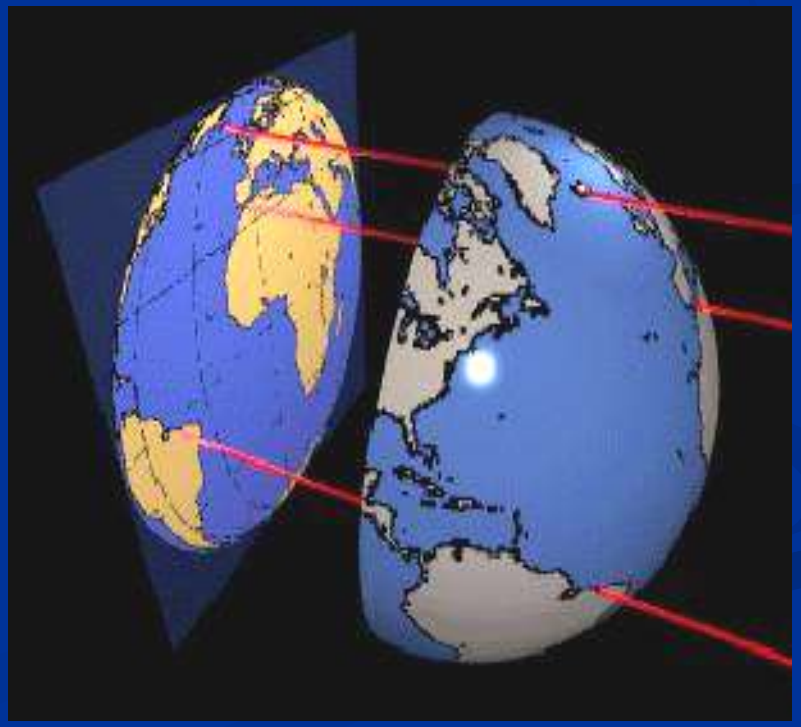
A Datum

- Uses a reference ellipsoid that conforms to the surface of the Geoid over the region of the planned mapping system



Map Projections

- The systematic transformation of points on the Earth's surface to corresponding points on a plane surface



Why Use Map Projections?

- It is much easier to measure distances on a plane
- The paper used to input and output GIS maps is flat
- Raster layers are made up of gridded cells which are flat
- The Earth has to be projected onto a flat surface to display it all at once

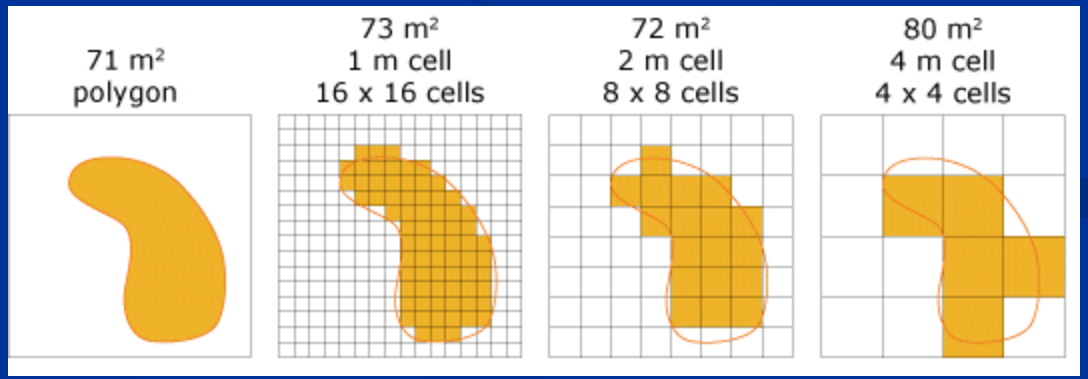
Key Issues in Managing Geospatial Data

- Scale and Resolution
- Accuracy and Precision
- Uncertainty in GIS
- Metadata

Scale and Resolution

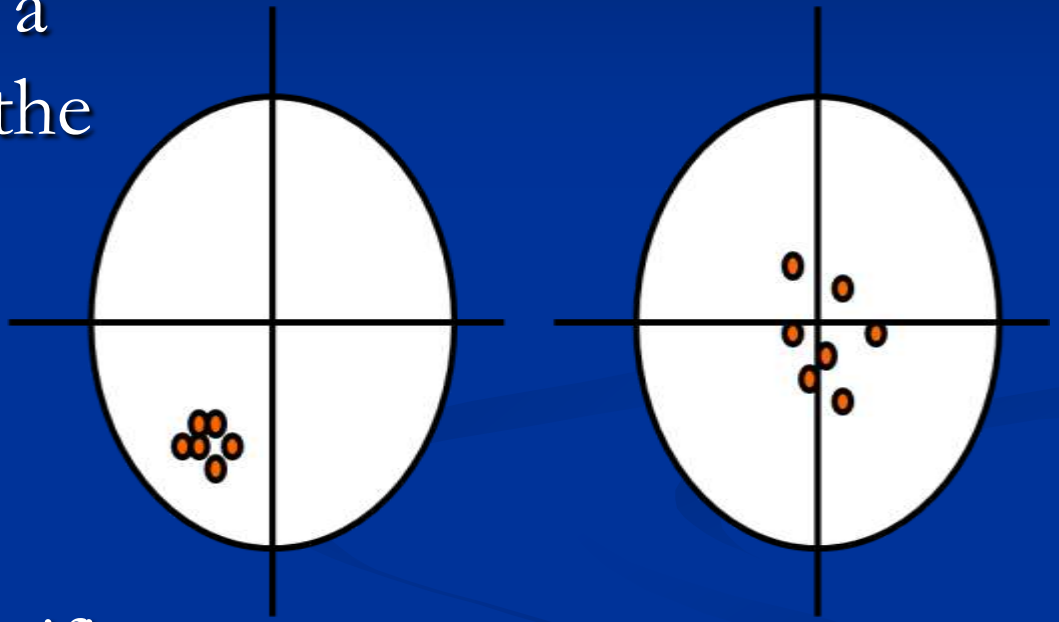
- *Scale:* how big ground objects are shown on a map
  - The ratio between distance on the map and distance on the ground
  - 1:1,250 large; 1:1,000,000 small

- *Resolution:* how much detail is identifiable on a map
  - *High*: more detail; smaller cell size



Precision and Accuracy

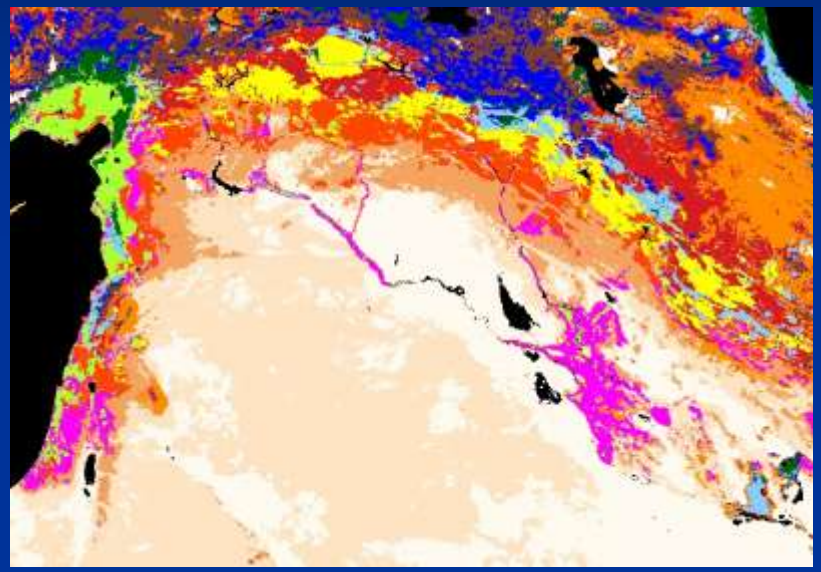
- *Accuracy*: how close a measurement is to the true value
- *Precision*:
  - Repeatability of measurements
  - The number of significant digits used to report a measurement



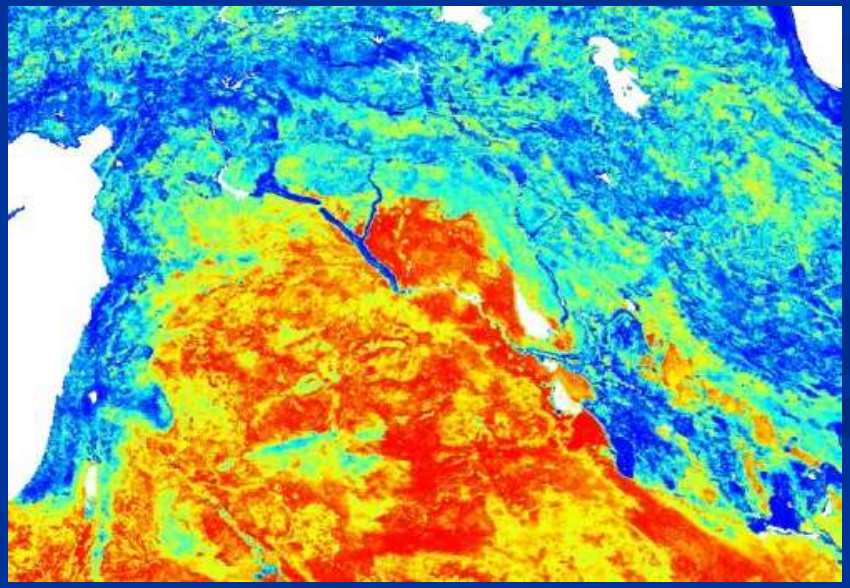
Uncertainty in GIS

- Accounts for the difference between the contents of a dataset and the phenomena that the data are supposed to represent
- Errors may occur in:
  - Positions
  - Attributes
  - Logic
  - Linage
  - Completeness, and
  - Time

Modeling Attribute Uncertainty



Uncertainty
Red: high
Blue: low

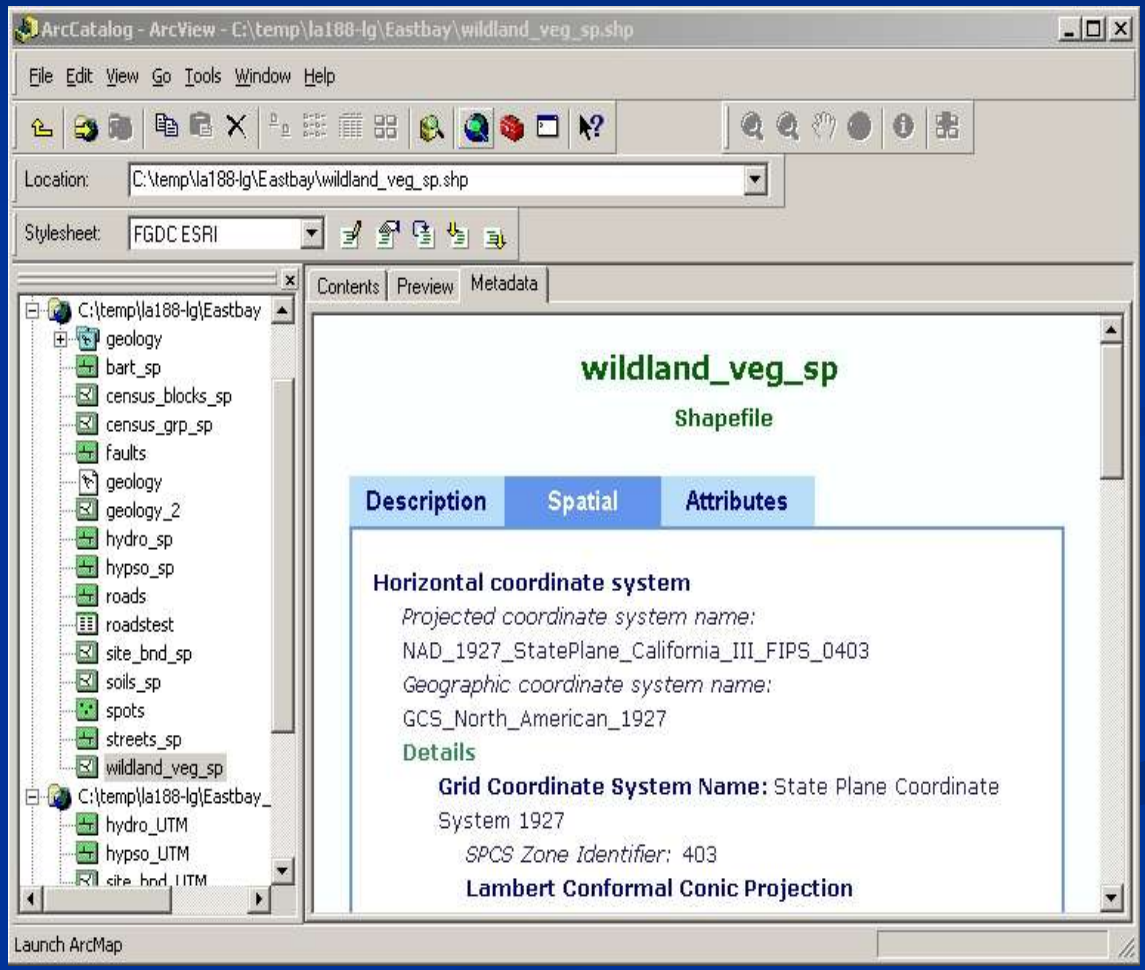


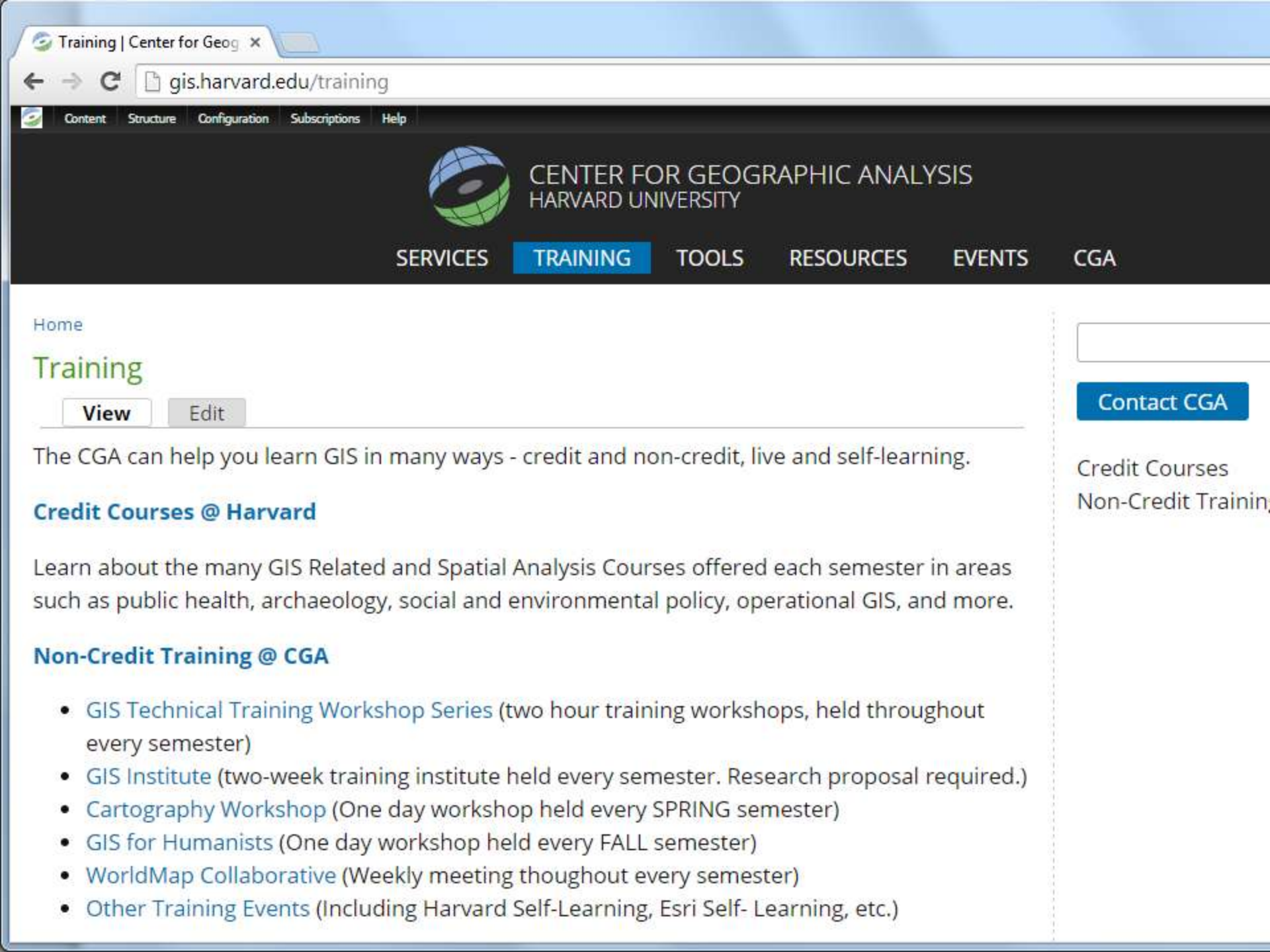
http://www.yale.edu/ceo/Projects/swap/landcover/uncertainty.htm

KCC

Metadata

- Data about data:
  - Identification
  - Data quality
  - Coordinate system
  - Attributes
  - …




Training | Center for Geog
gis.harvard.edu/training
Content
Structure
Configuration
Subscriptions
Help
CENTER FOR GEOGRAPHIC ANALYSIS
HARVARD UNIVERSITY
SERVICES
TRAINING
TOOLS
RESOURCES
EVENTS
CGA
Home
Training
View
Edit
The CGA can help you learn GIS in many ways - credit and non-credit, live and self-learning.
Credit Courses @ Harvard
Learn about the many GIS Related and Spatial Analysis Courses offered each semester in areas such as public health, archaeology, social and environmental policy, operational GIS, and more.
Non-Credit Training @ CGA
GIS Technical Training Workshop Series (two hour training workshops, held throughout every semester)
GIS Institute (two-week training institute held every semester. Research proposal required.)
Cartography Workshop (One day workshop held every SPRING semester)
GIS for Humanists (One day workshop held every FALL semester)
WorldMap Collaborative (Weekly meeting thoughout every semester)
Other Training Events (Including Harvard Self-Learning, Esri Self- Learning, etc.)
Contact CGA
Credit Courses
Non-Credit Trainin

References

- Longley, P. A., M. F. Goodchild, D. J. Maguire, D. W. Rhind (2010). Geographic Information Systems & Science (3rd Ed). John Wiley & Sons, Inc.
- Clarke, K.C. (2010). Getting Started with GIS (5th Ed). Prentice-Hall, Inc., London.
- Zeiler, M. (1999). Modeling Our World: The ESRI Guide to Geodatabase Design. ESRI Press, California.
- Chang, K-T. (2010). Introduction to Geographic Information Systems (5th Ed). McGraw-Hill. (KTC)
- Center for Geographic Analysis (CGA) Harvard: GIS Training Materials
- ESRI: Introduction to ARC/INFO Training Materials