Spatial data in R

Matthew L. Sisk
Associate Professor of the Practice
Lucy Family Institute for Data & Society

https://github.com/msisk1/workshops/tree/master/SpatialR

Workshop overview

We will follow part of the material for the Carpentries R for GIS curriculum

https://datacarpentry.org/r-raster-vector-geospatial/

The Carpentries uses a hands-on, code along approach (as opposed to lecture or demonstration).

Odds of running into an error are high, but don't panic, we will all encounter issues (even the instructors!) and we will work through them together!

What is GIS ?

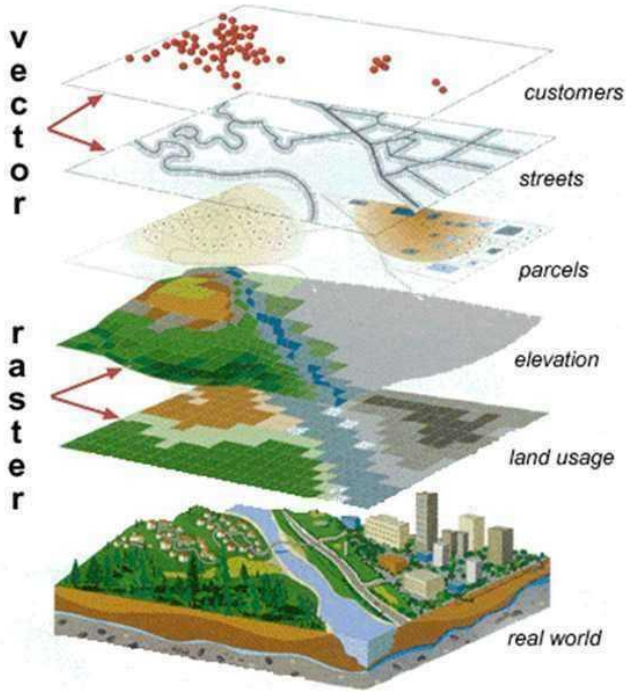
In short: “computerized mapping software”

Formal definition

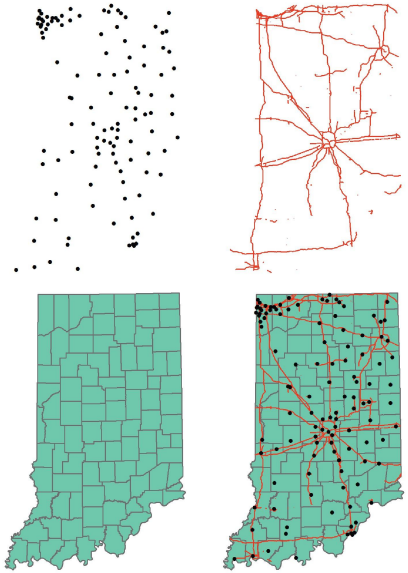
A Geographic Information System (GIS) is a computerized database management system for capture, storage, retrieval, manipulation, analysis and display of spatial (i.e. locationally defined) data

Layers

- A GIS is composed of layers of spatial information
- Can be different types of data
- Everything is referenced to a coordinate system
  - e.g. latitude / longitude

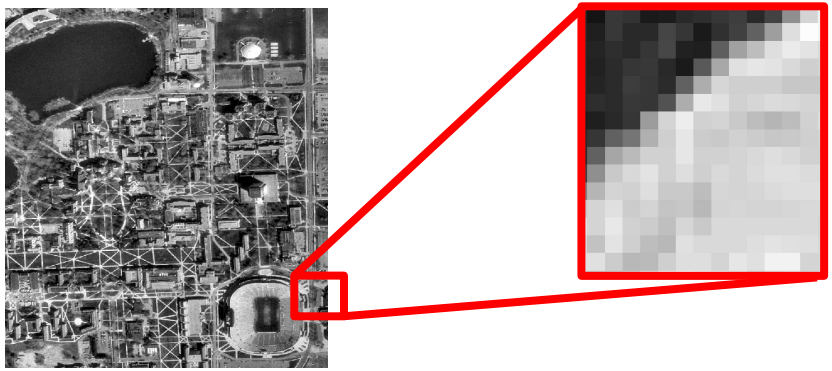


GIS digitally models the real world using:



Three types of geometry

- -Points
- -Lines
- -Areas



Cells in an image

Table

Indiana Counties

FID	Shape *	NAME_L	POP2000	MALES	FEMALES	AGE_UNDER5	AGE_5_17	AGE_18_21	AGE_22_29
0	Polygon	Steuben	33214	16771	16443	2199	6322	2241	3307
1	Polygon	Lagrange	34909	17681	17228	3432	8381	2199	3674
2	Polygon	Elkhart	182791	90848	91943	14800	37999	9881	20712
3	Polygon	St Joseph	265559	128133	137426	18673	49616	20658	28143
4	Polygon	Lake	484564	233367	251197	34639	95158	26621	48719
5	Polygon	Porter	146798	72046	74752	9488	28314	9093	14112
6	Polygon	La Porte	110106	56539	53567	7116	19886	5454	11258
7	Polygon	De Kalb	40285	20059	20226	3061	8238	1977	4210
8	Polygon	Noble	46275	23310	22965	3695	9729	2441	5074
9	Polygon	Marshall	45128	22415	22713	3290	9369	2335	4256
10	Polygon	Kosciusko	74057	36982	37075	5519	15043	3728	7655
11	Polygon	Starke	23556	11660	11896	1520	4792	1142	2169
12	Polygon	Whitley	30707	15238	15469	2101	6112	1469	2853
13	Polygon	Allen	331849	162425	169424	25440	66511	18022	36702
14	Polygon	Jasper	30043	14888	15155	2077	6157	1985	2856
15	Polygon	Newton	14566	7239	7327	902	2945	728	1265
16	Polygon	Fulton	20511	10139	10372	1348	3986	961	1835
17	Polygon	Pulaski	13755	6938	6817	845	2858	610	1171
18	Polygon	Wabash	34960	16957	18003	2073	6504	2377	3234
19	Polygon	Huntington	38075	18537	19538	2536	7412	2395	3600

0 (0 out of 92 Selected)

Indiana Counties

Data tables

GIS and GPS

GPS: Global Positioning System

- A system of 24 satellites that enabled devices can use to get a precise location on the globe
- GIS frequently uses data from GPS receivers but the two are not linked

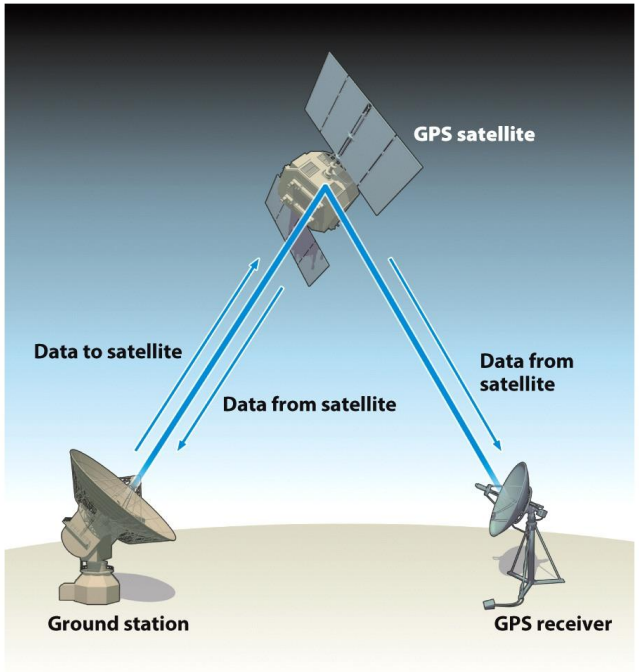
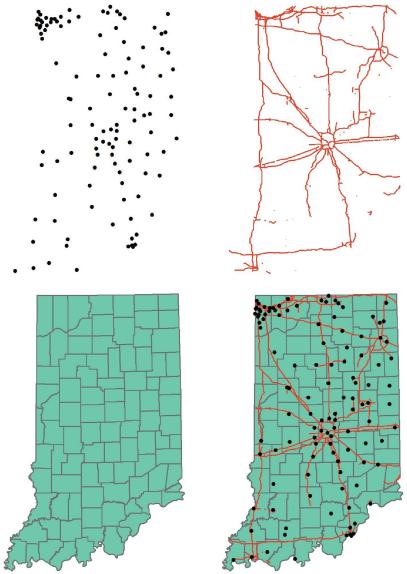
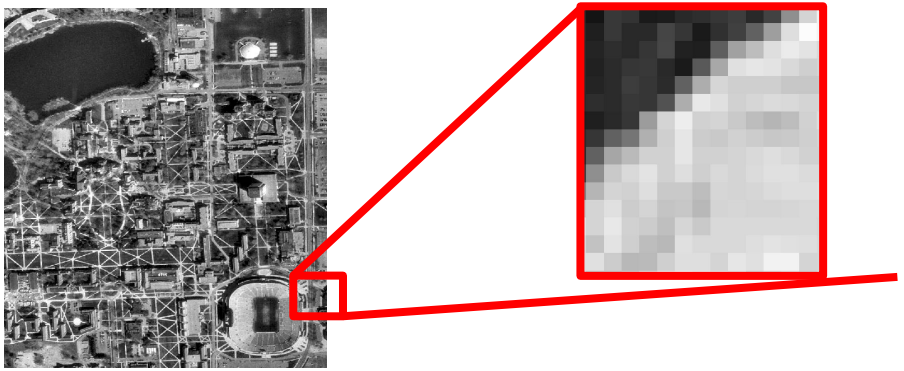


Figure 4.2
Introduction to Geospatial Technologies, Second Edition



Three types of geometry

- -Points
- -Lines
- -Polygons



Cells in an image

Table

Indiana Counties

FID	Shape *	NAME_L	POP2000	MALES	FEMALES	AGE_UNDER5	AGE_5_17	AGE_18_21	AGE_22_29
0	Polygon	Steuben	33214	16771	16443	2199	6322	2241	3307
1	Polygon	Lagrange	34909	17681	17228	3432	8381	2199	3674
2	Polygon	Elkhart	182791	90848	91943	14800	37999	9881	20712
3	Polygon	St Joseph	265559	128133	137426	18673	49616	20658	28143
4	Polygon	Lake	484564	233367	251197	34639	95158	26621	48719
5	Polygon	Porter	146798	72046	74752	9488	28314	9093	14112
6	Polygon	La Porte	110106	56539	53567	7116	19886	5454	11258
7	Polygon	De Kalb	40285	20059	20226	3061	8238	1977	4210
8	Polygon	Noble	46275	23310	22965	3695	9729	2441	5074
9	Polygon	Marshall	45128	22415	22713	3290	9369	2335	4256
10	Polygon	Kosciusko	74057	36982	37075	5519	15043	3728	7655
11	Polygon	Starke	23556	11660	11896	1520	4792	1142	2169
12	Polygon	Whitley	30707	15238	15469	2101	6112	1469	2853
13	Polygon	Allen	331849	162425	169424	25440	66511	18022	36702
14	Polygon	Jasper	30043	14888	15155	2077	6157	1985	2856
15	Polygon	Newton	14566	7239	7327	902	2945	728	1265
16	Polygon	Fulton	20511	10139	10372	1348	3986	961	1835
17	Polygon	Pulaski	13755	6938	6817	845	2858	610	1171
18	Polygon	Wabash	34960	16957	18003	2073	6504	2377	3234
19	Polygon	Huntington	38075	18537	19538	2536	7412	2395	3600

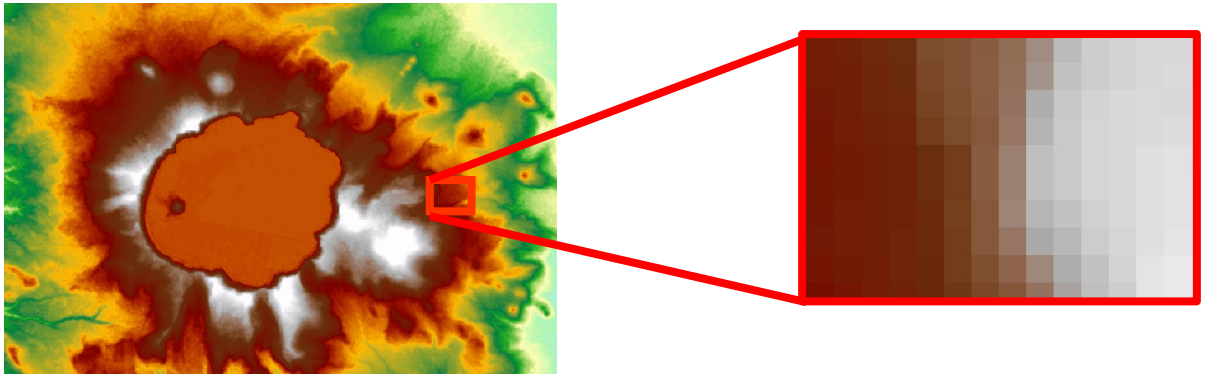
0 (0 out of 92 Selected)

Indiana Counties

Data tables

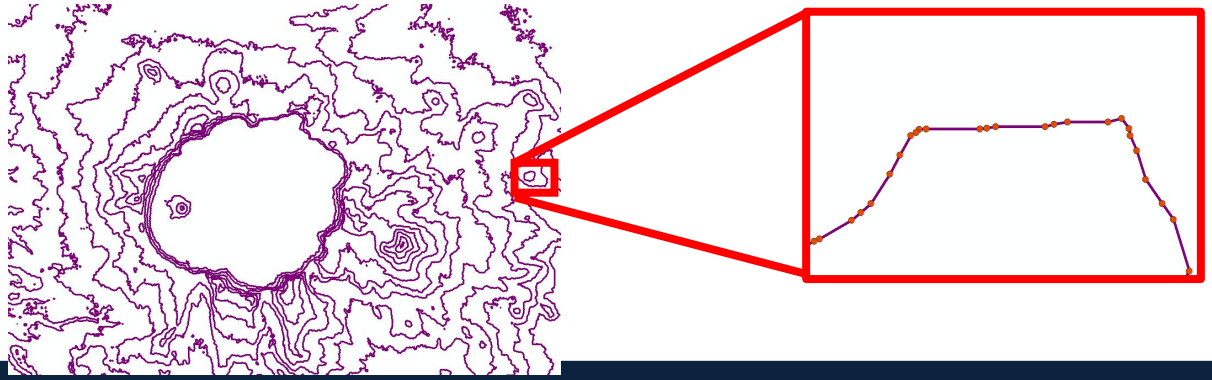
Raster Data

Based on pixel



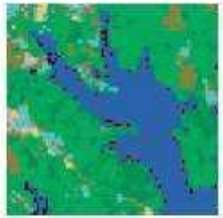
Vector Data

Based on discrete points

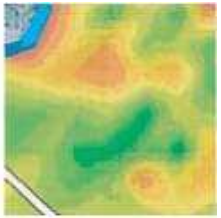


Rasters

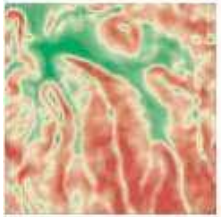
Orthophoto



Land Use

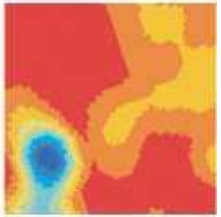


Concentration



Slope

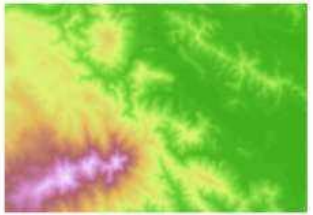
Elevation



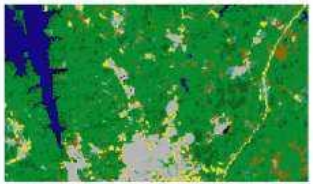
Population

Aerial Imagery

Surface Elevation



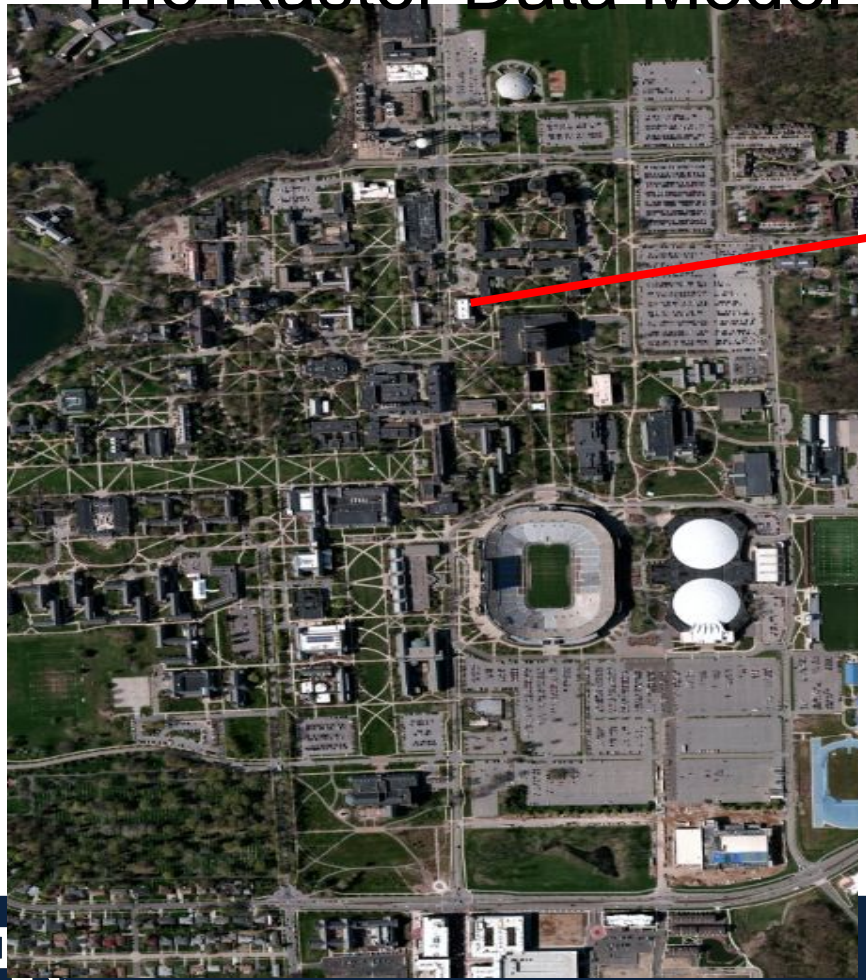
Land Use Classes

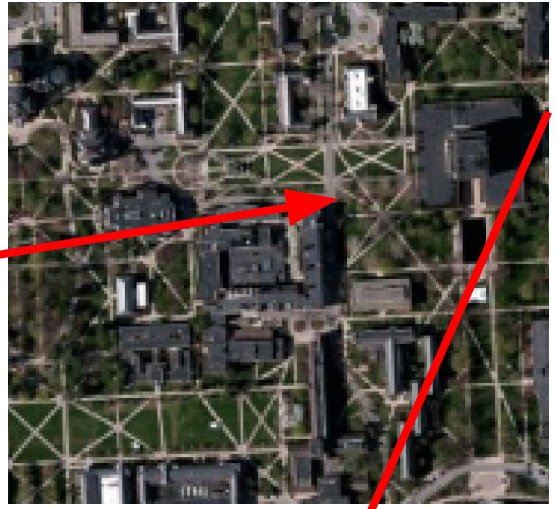


- Agriculture
- Bare ground
- Water
- Deciduous Trees
- Deciduous / Pine Mixed
- Grassland
- Pine
- Shadow
- Urban / Developed

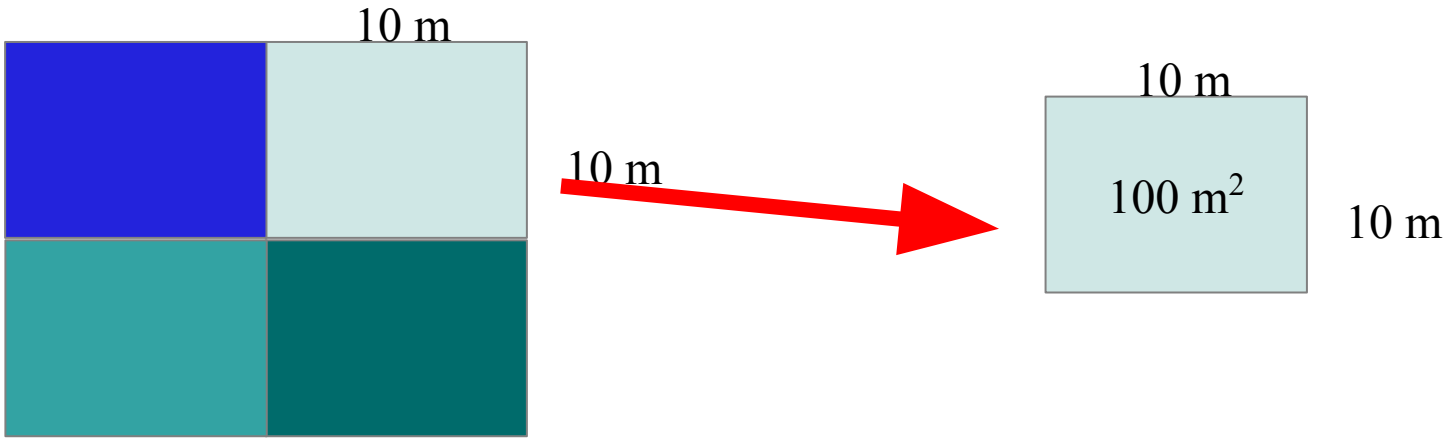
- The fundamental unit of a raster image is the pixel
- This is the same as a digital picture

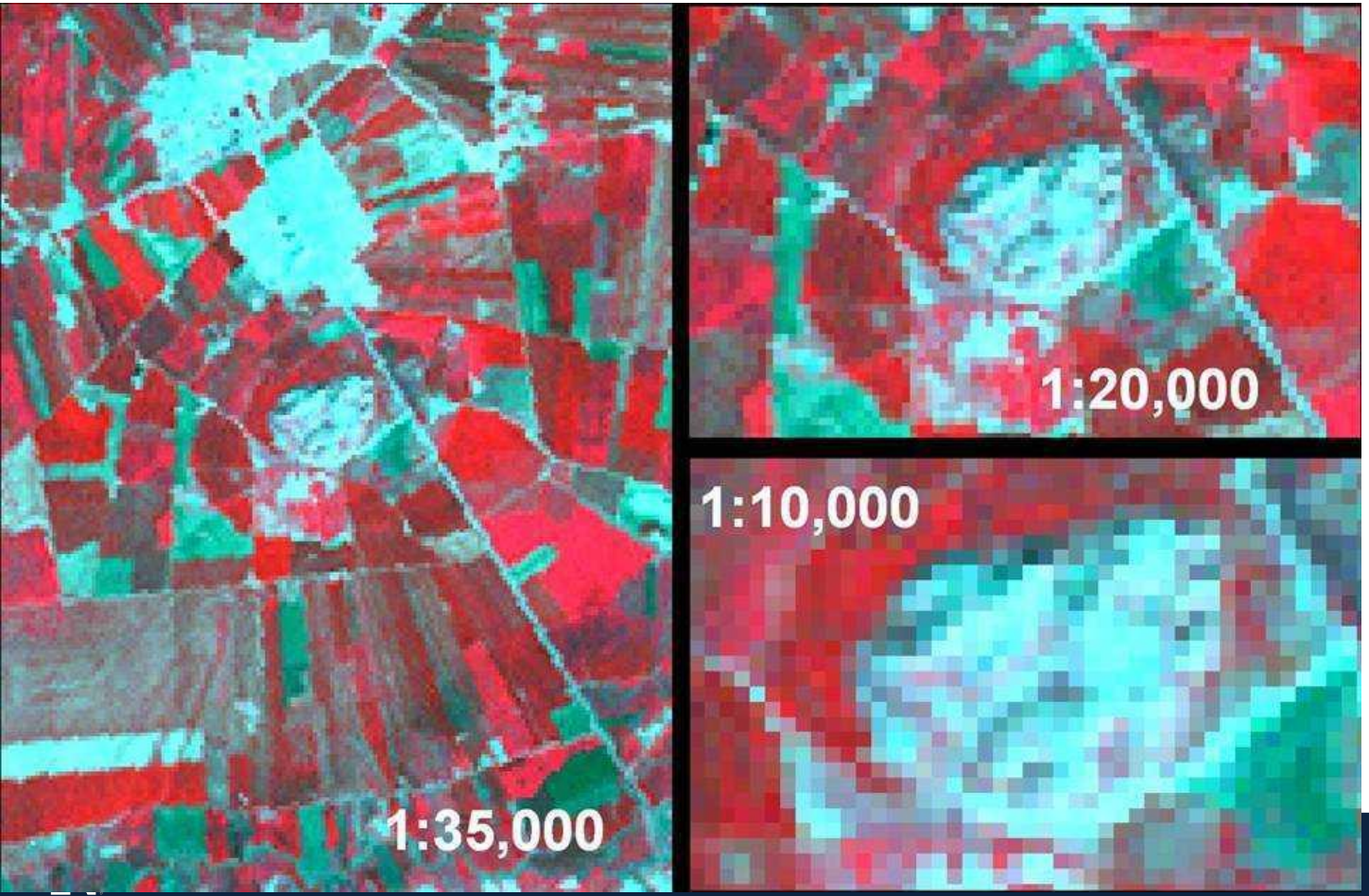
The Raster Data Model





The length, in real world measurements, of each side of a square pixel




1:35,000
1:20,000
1:10,000

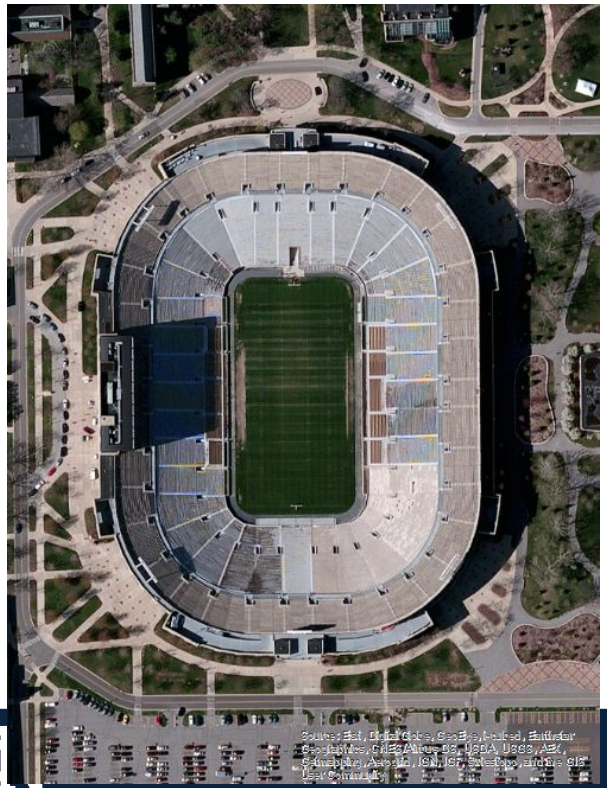
Raster Advantages

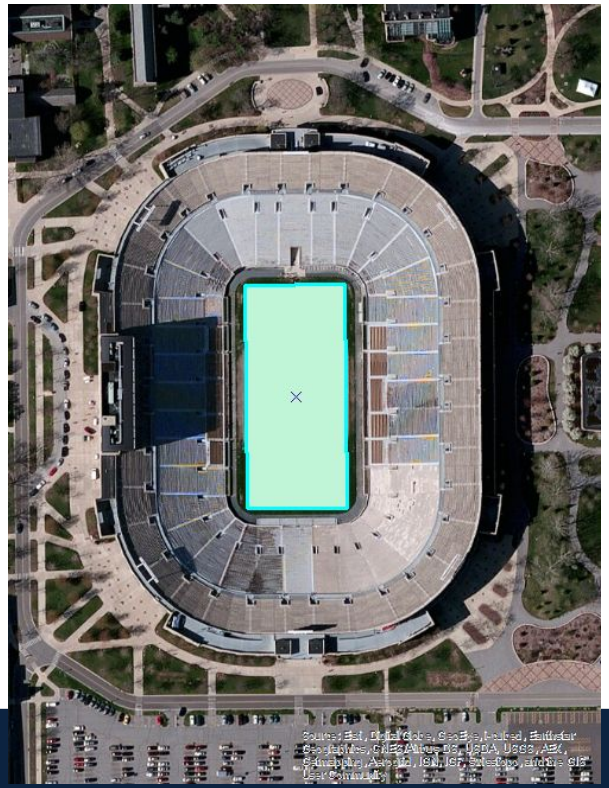
- Continuous Coverage
- Detail beyond human perception
- Easily manipulated

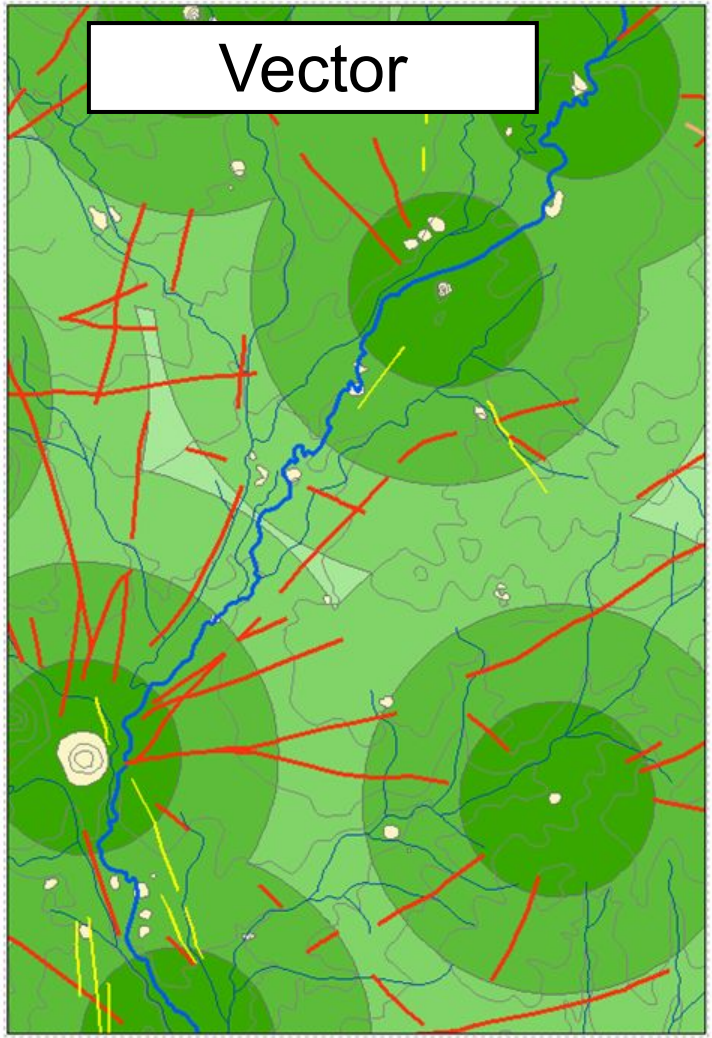
Raster Disadvantages

- Fixed resolution
- Large file size
- Difficult to edit individual pixels

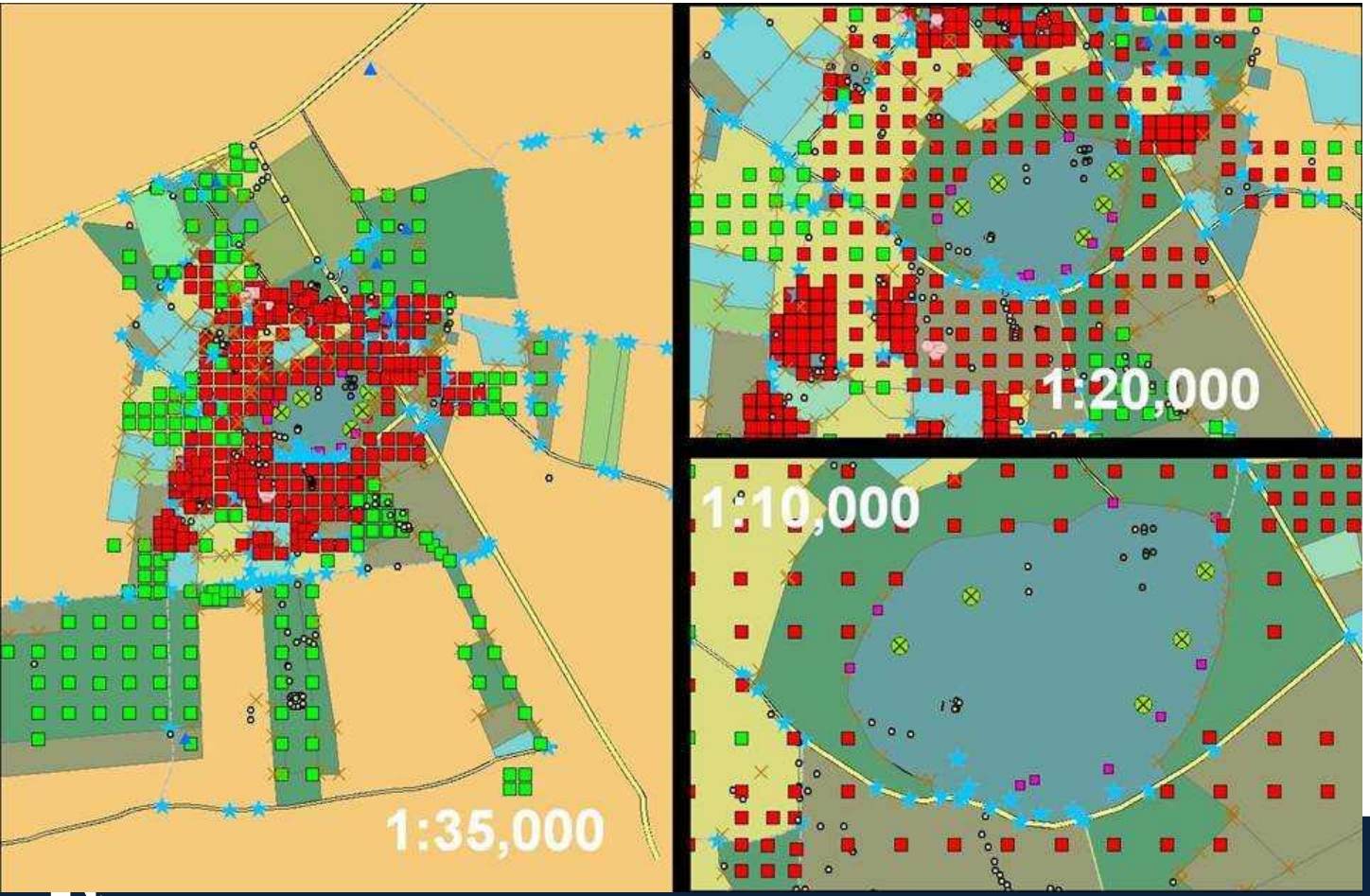
Raster and Vector






Vector


Raster
0
1
2
3
Km


1:35,000
1:20,000
1:10,000

Vector Advantages

- Economical in space
- Good for discrete features
- More flexible with regard to scale

Vector Disadvantages

- More schematized version of reality
- Poorly suited for continuous phenomena

Location

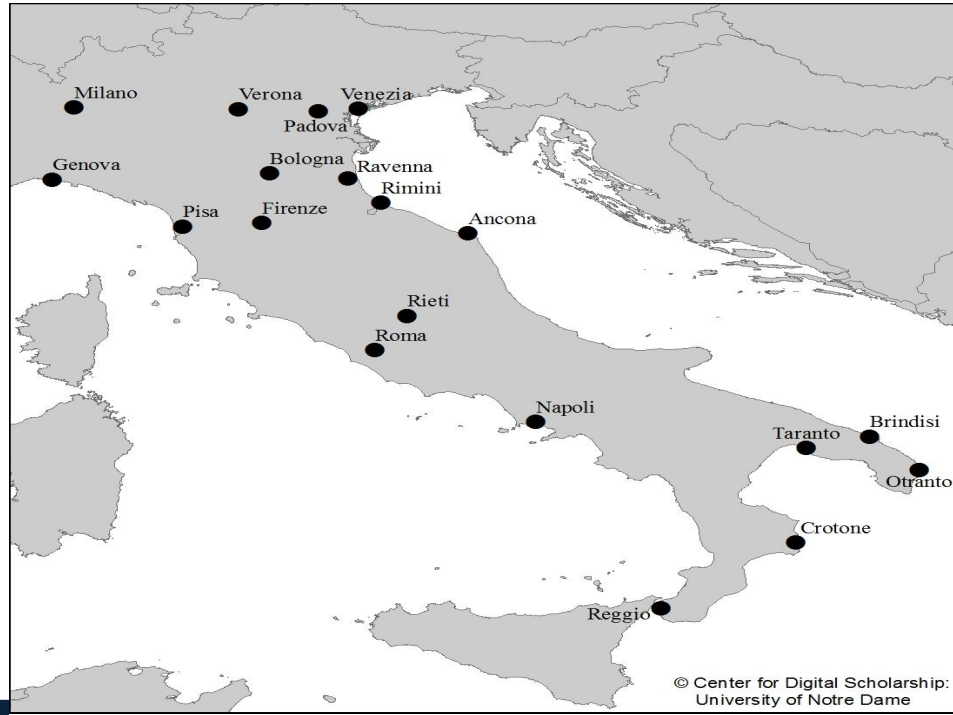
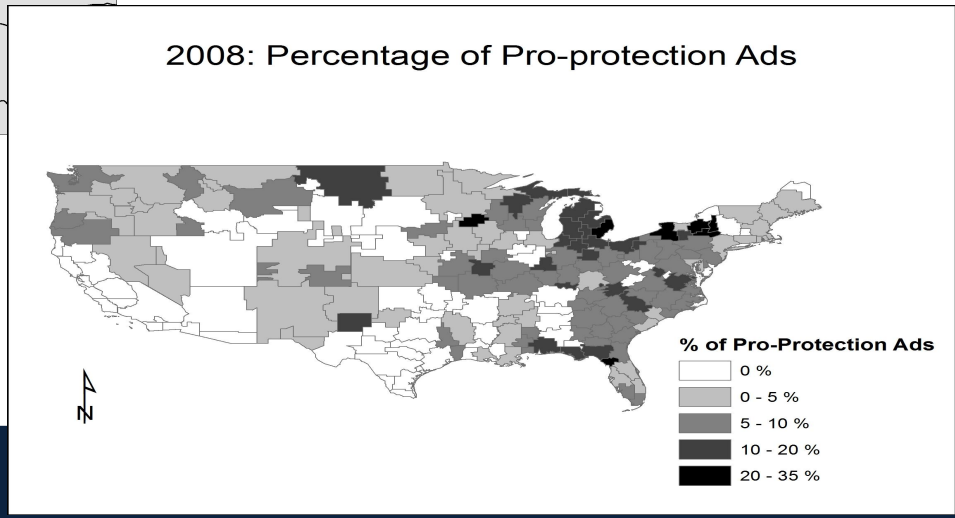


Illustration 3: La mappa dell'Italia di Petrarca, Epistola metrica, 2.11

Classification



Example: Patterns

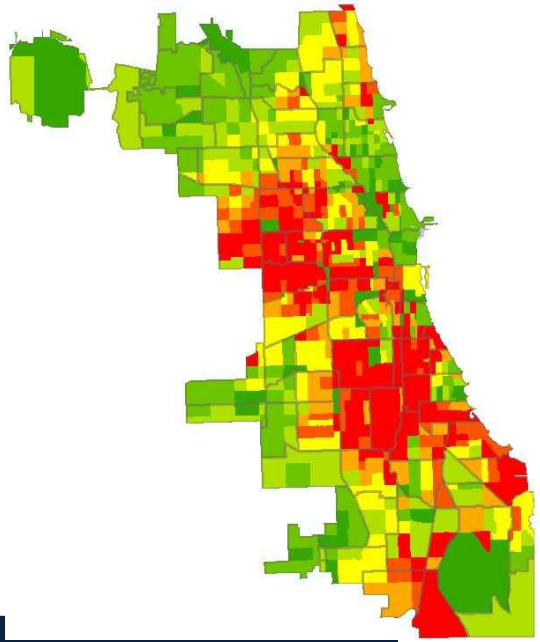
Socio-economic

2000 Census data

Unit – Census Tract

Over the community area boundaries

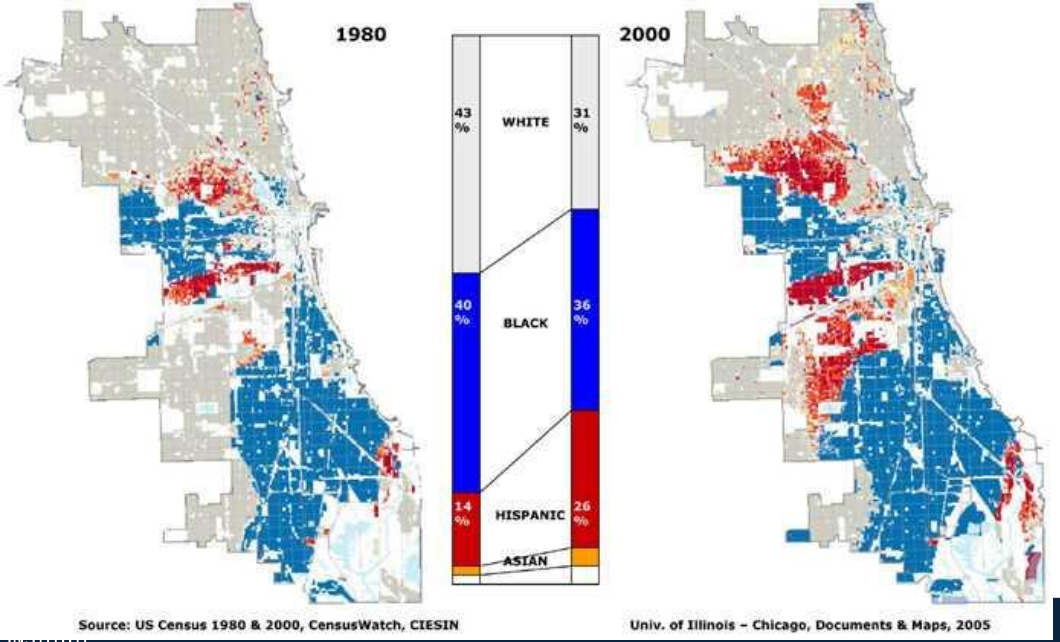
% families below poverty level



Example: Trends

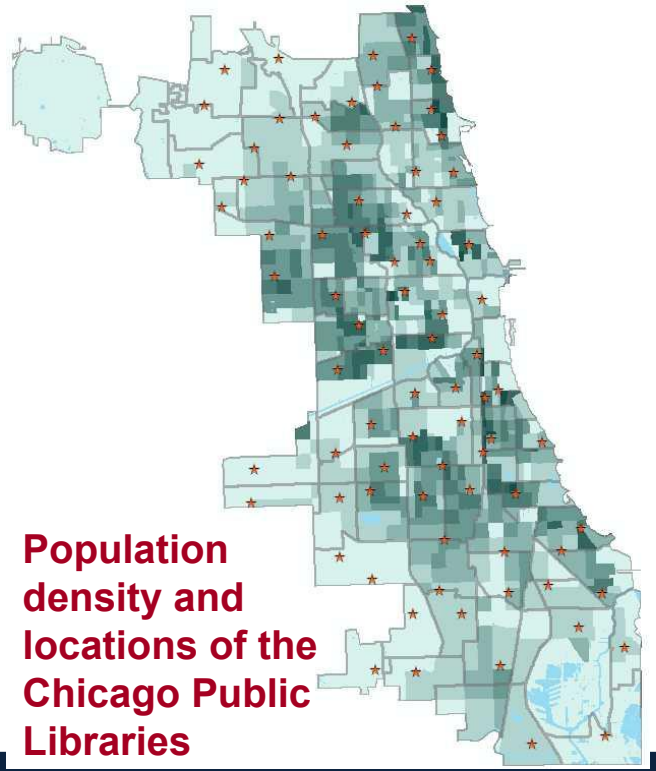
Trends

Changes over time using historical (time-series) data



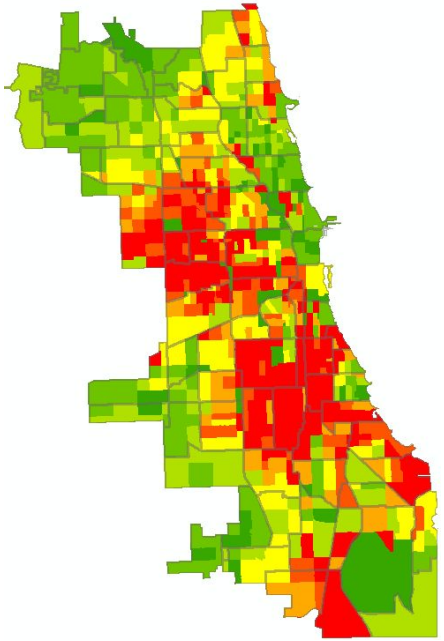
Example: Relationship

- Association
  - Demography and Libraries
  - Are there enough libraries to serve local residents?

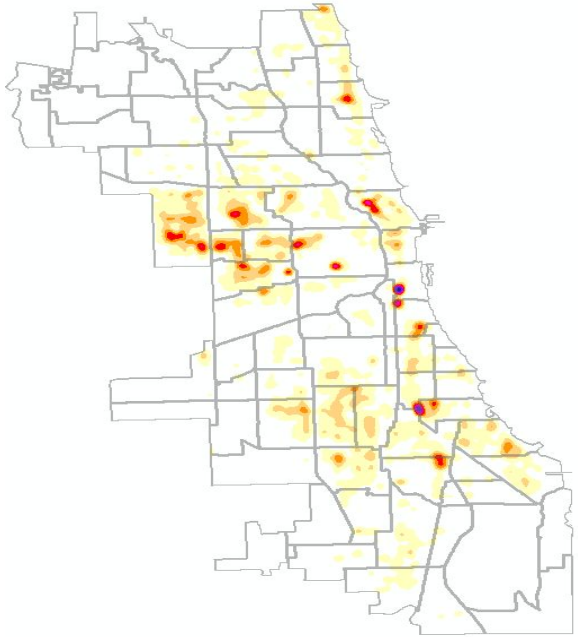


Population density and locations of the Chicago Public Libraries

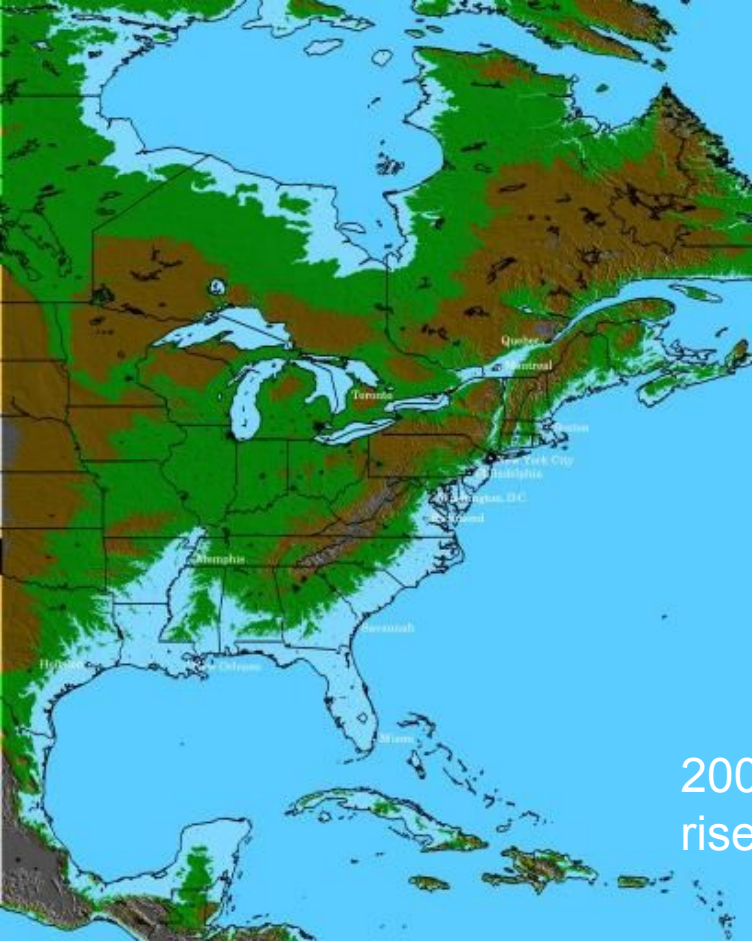
Example: Associations



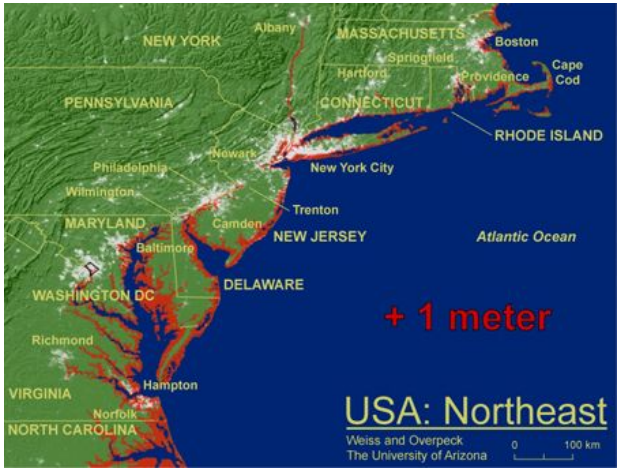
% families below poverty level



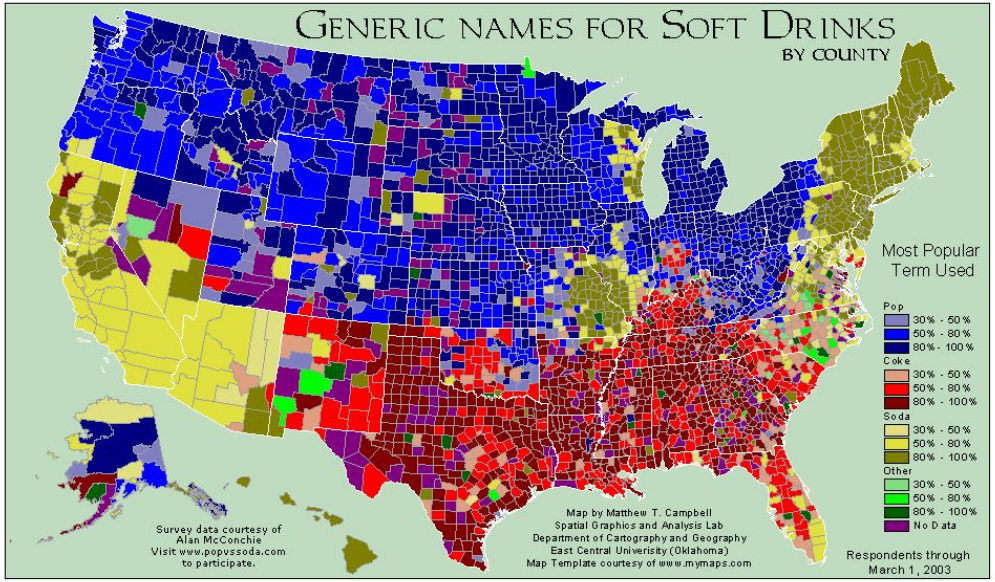
Crime hotspots



200
rise



Areas impacted by sea-level rise

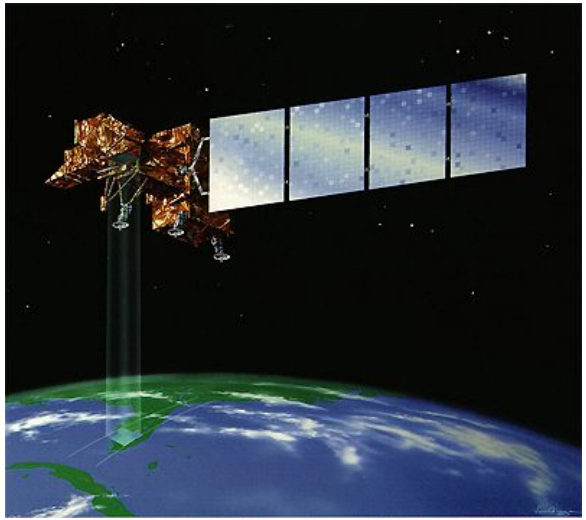


Geographers have even analyzed the distribution of generic names for soft drinks by US

Same analysis, but done via geotagged tweets mentioning one of the key words

Satellite Imagery

- **Remote sensing** is the science of acquiring, processing and interpreting information/data collected by remote sensors.
- Often, these are Earth Observing Satellites
  - Record visible and non-visible wavelengths of light



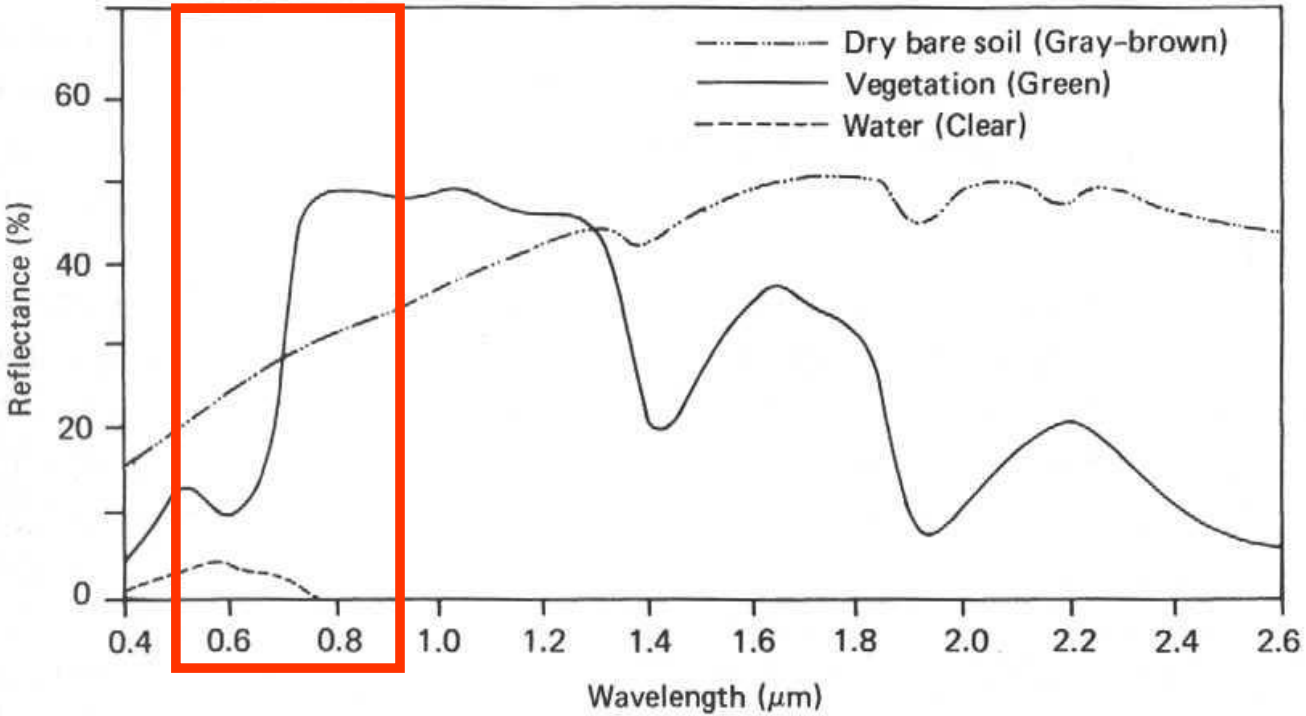


Figure 1.10 Typical spectral reflectance curves for vegetation, soil, and water. (Adapted from Swain and Davis, 1978.)

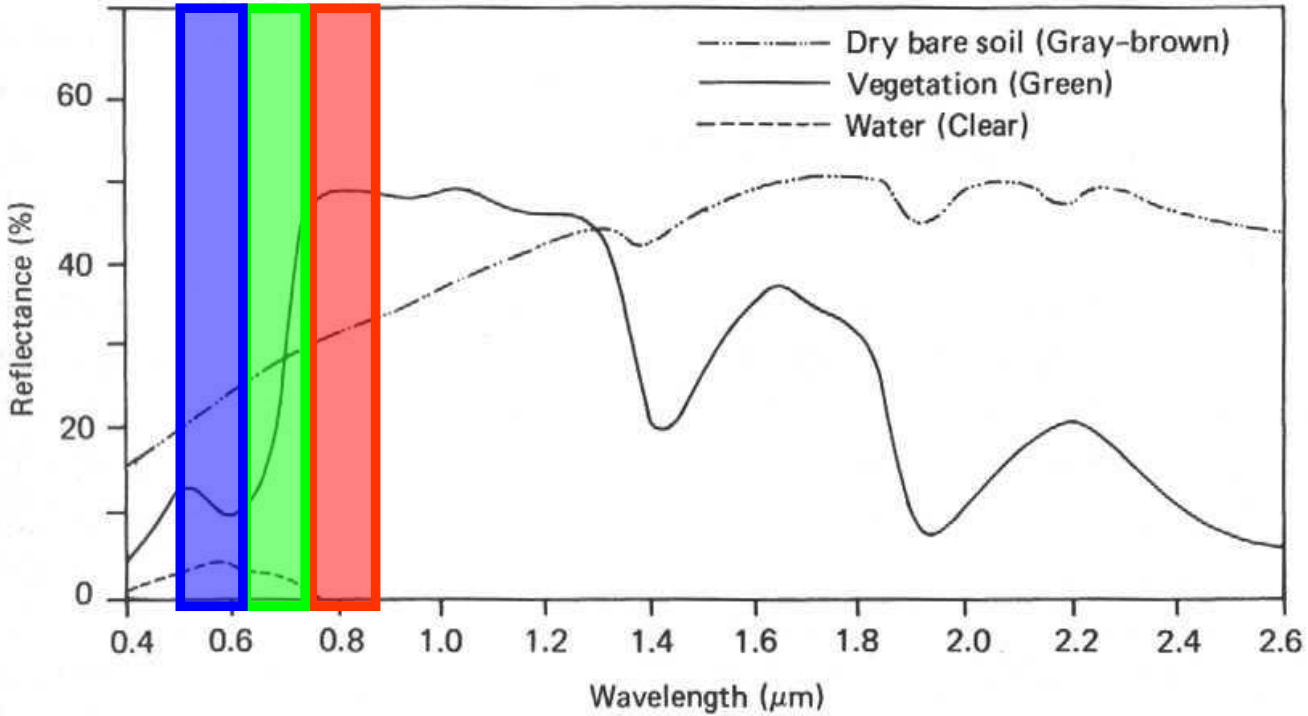
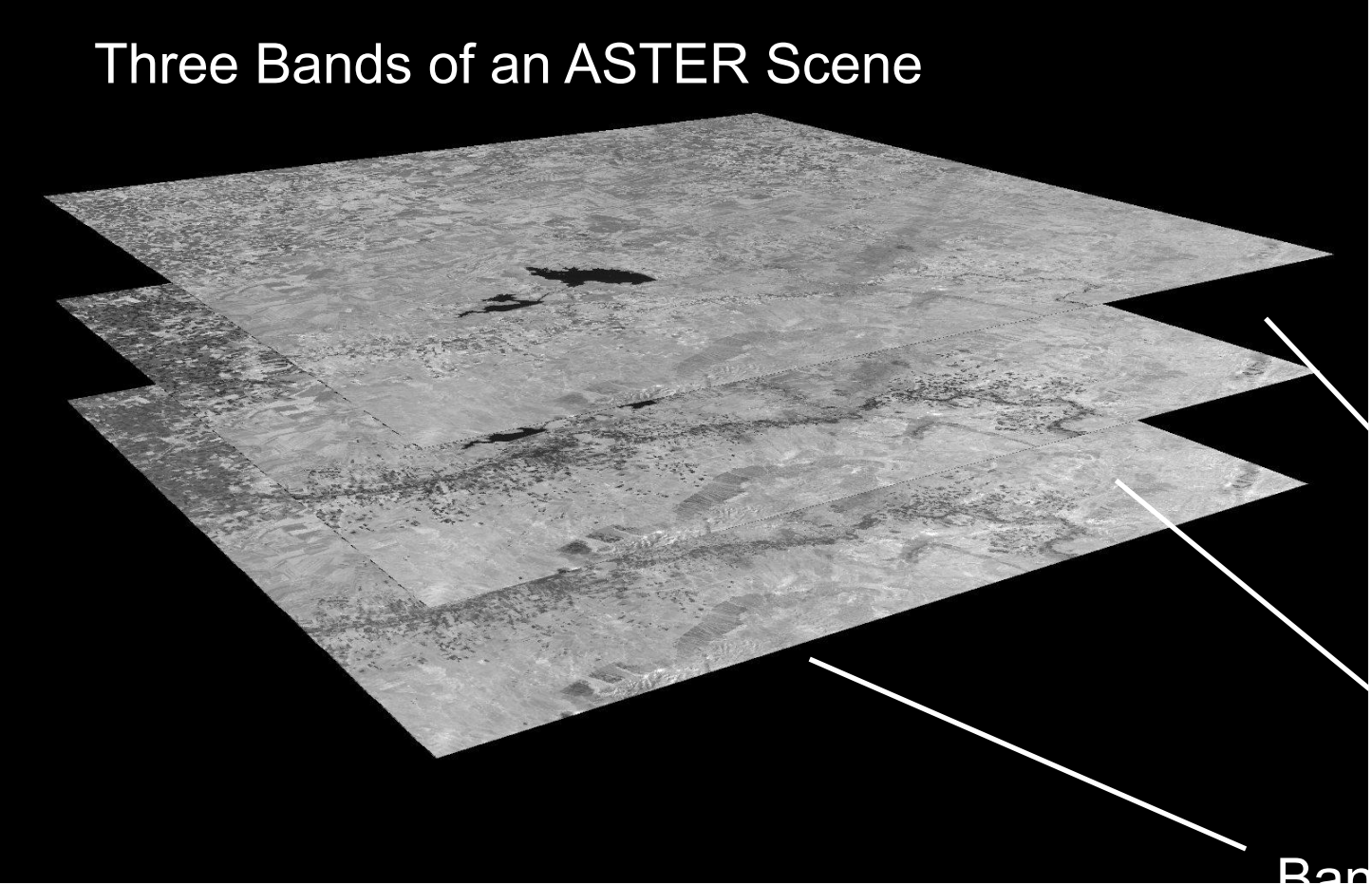
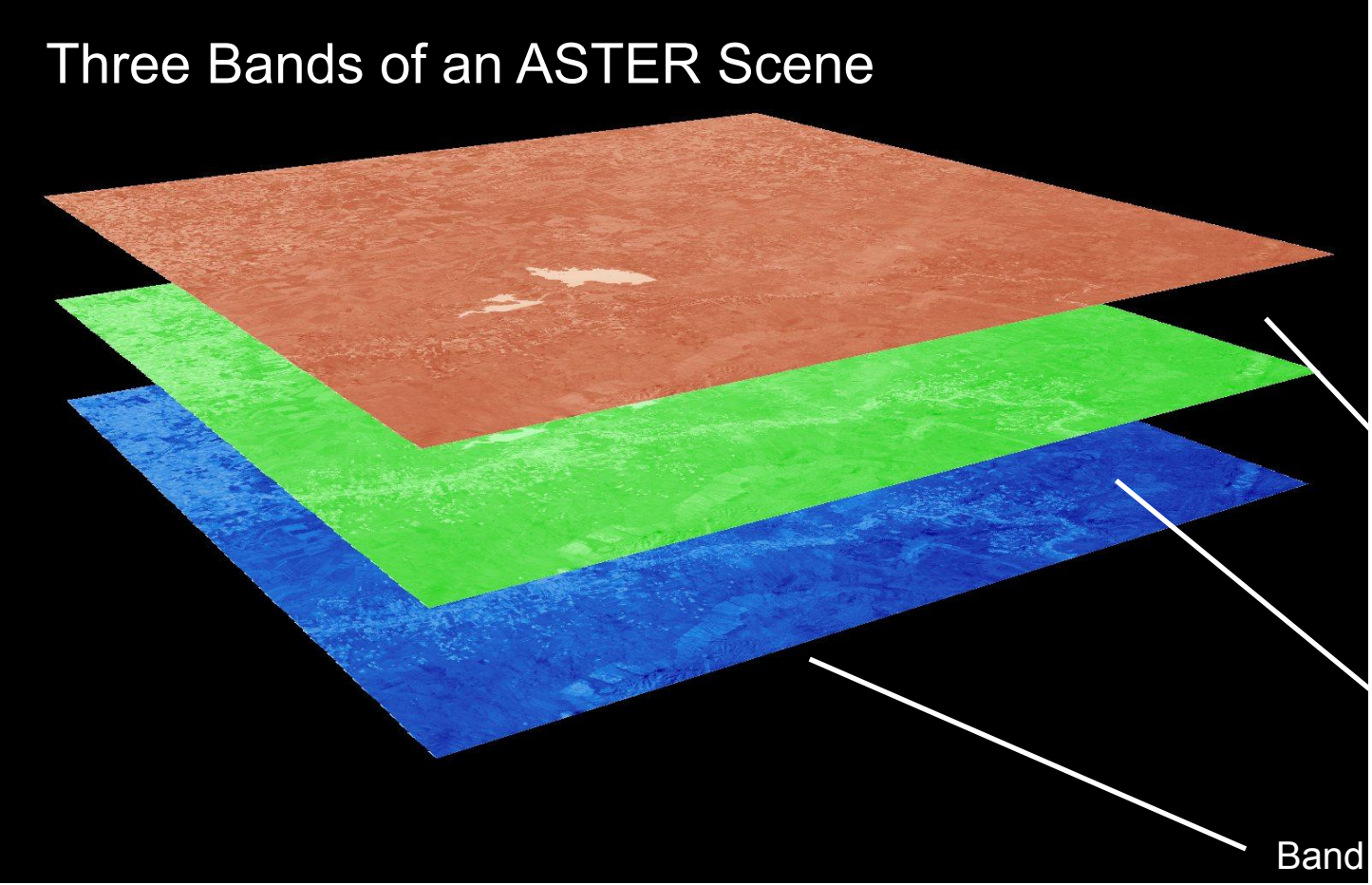


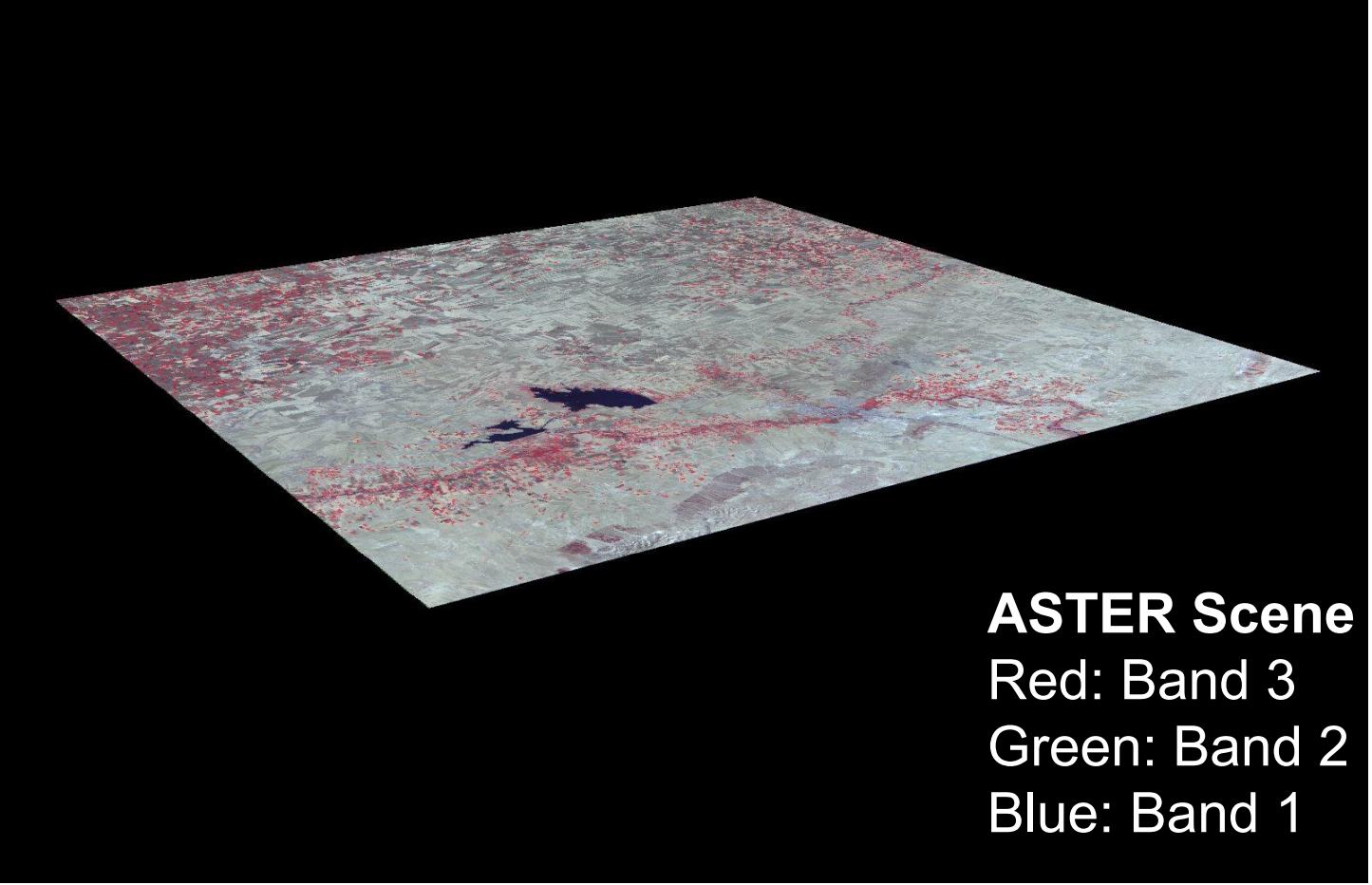
Figure 1.10 Typical spectral reflectance curves for vegetation, soil, and water. (Adapted from Swain and Davis, 1978.)

Three Bands of an ASTER Scene



Three Bands of an ASTER Scene





ASTER Scene
Red: Band 3
Green: Band 2
Blue: Band 1

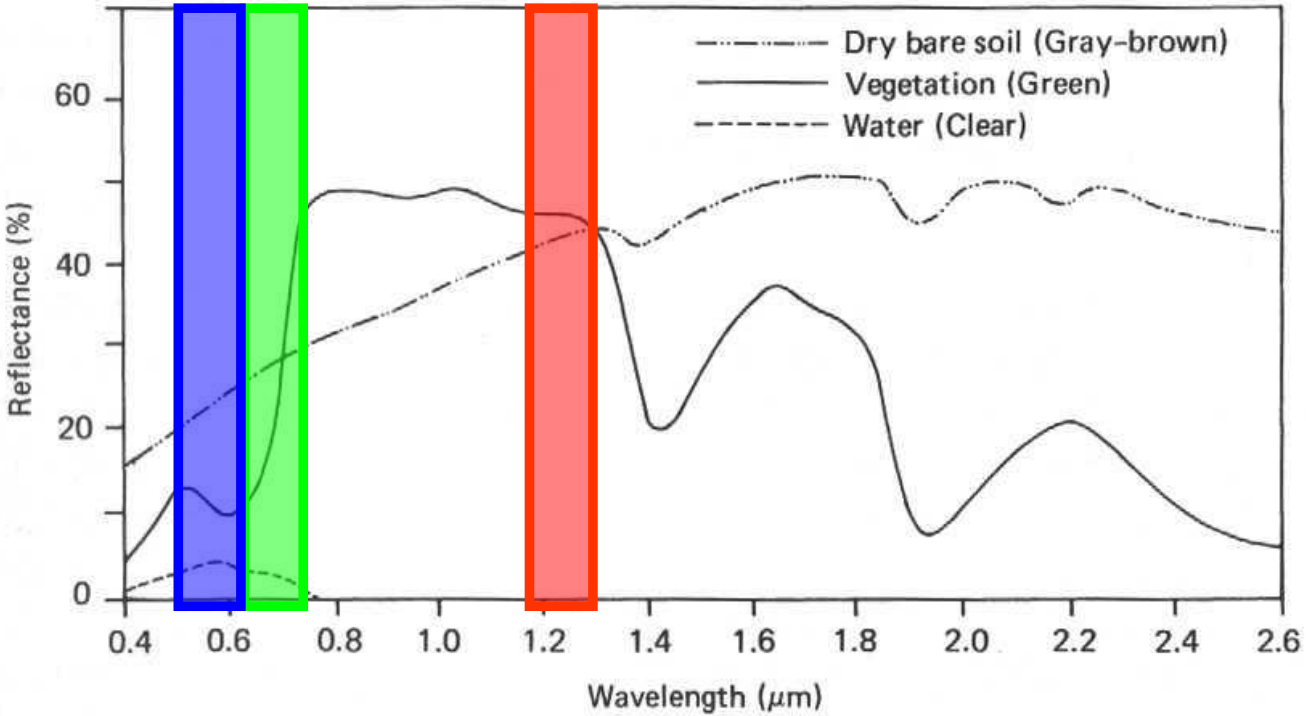
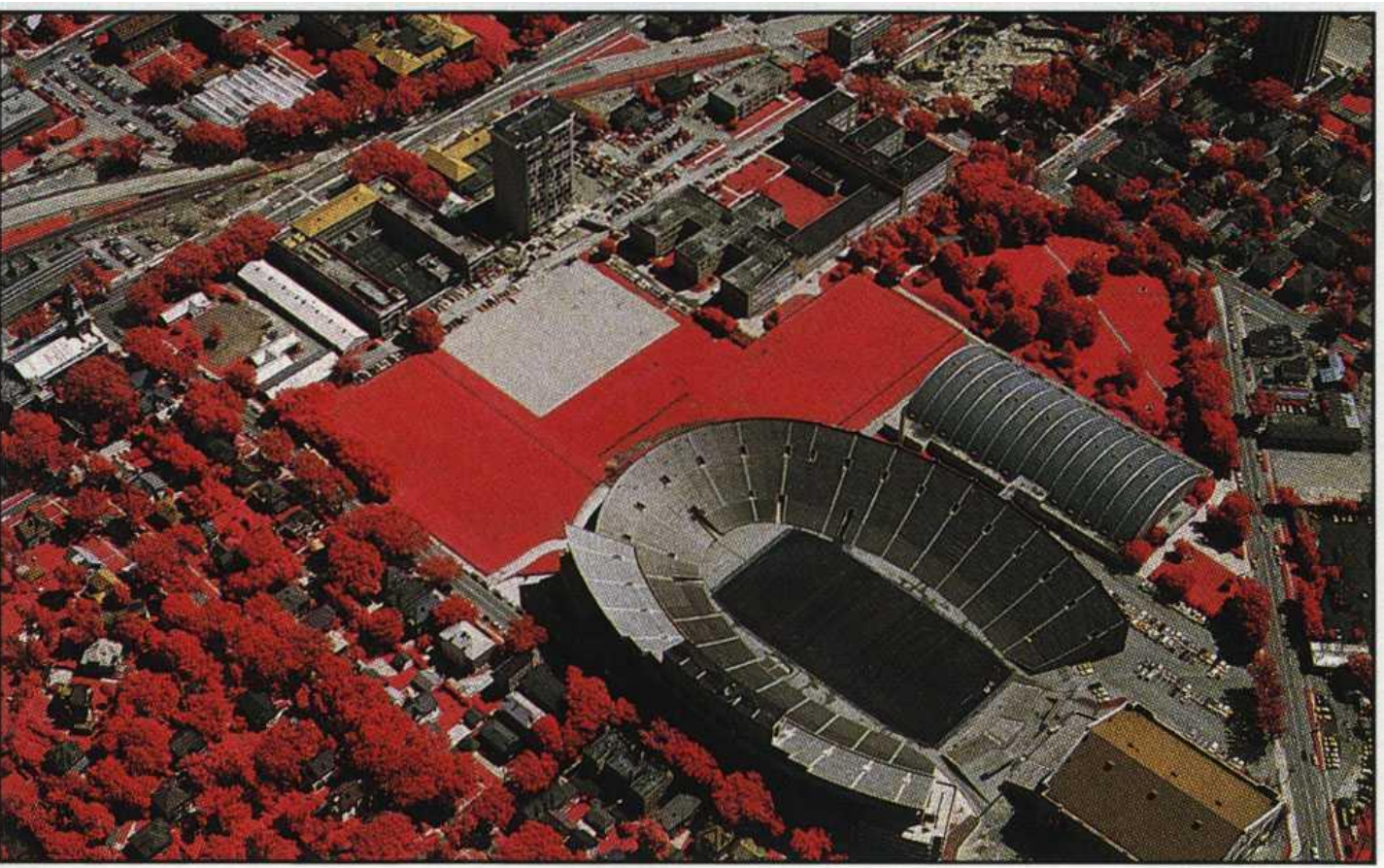
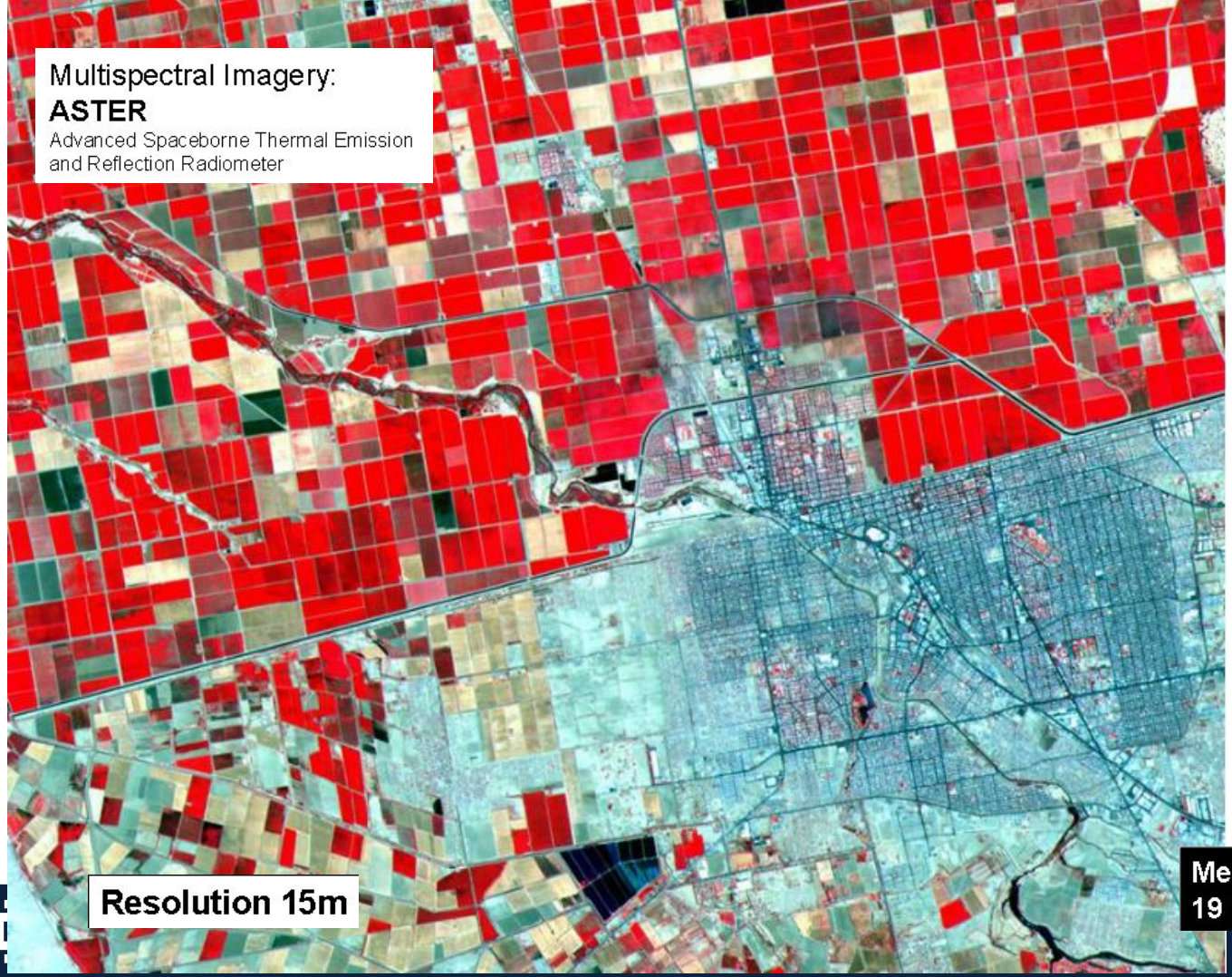
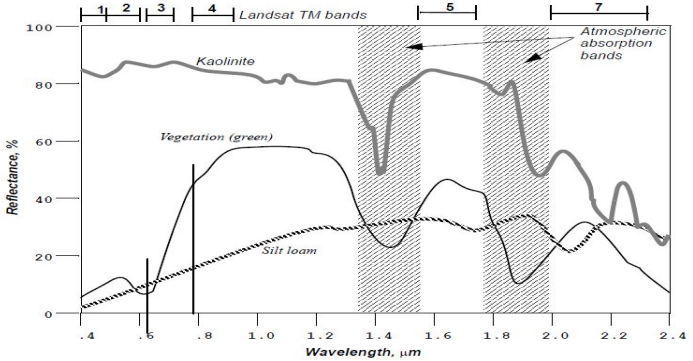


Figure 1.10 Typical spectral reflectance curves for vegetation, soil, and water. (Adapted from Swain and Davis, 1978.)




Multispectral Imagery:
ASTER
Advanced Spaceborne Thermal Emission and Reflection Radiometer
Resolution 15m
Mexicali, California
19 May 2000

Vegetation Indices



Based on the ratio of Near Infrared to Red
- **Chlorophyll reflects strongly in Infrared**
- **Red measure removes some error**

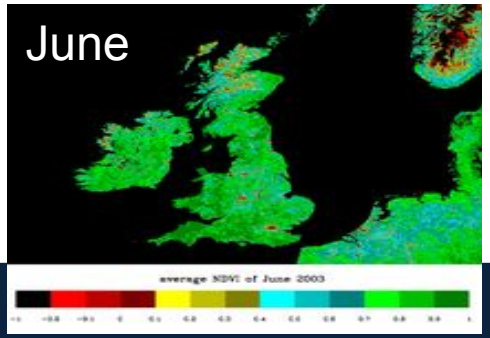
Measures vegetation density and health

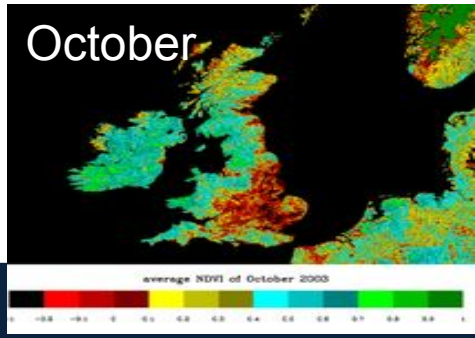
Several possible measures

Normalized Difference Vegetation Index (NDVI)

NDVI values range between -1 and 1
- 1 – 0.1 : Clouds, water, rock, sand, snow.
- 0.2 – 0.4 : Shrub and grasslands
- 0.5 – 0.8 : Dense forest

In ArcGIS NVDI is scaled to 0 – 200:
- 0 – 110 : Not vegetation
- 120 – 140 : Shrub and grasslands
- 150 – 180 : Dense forest

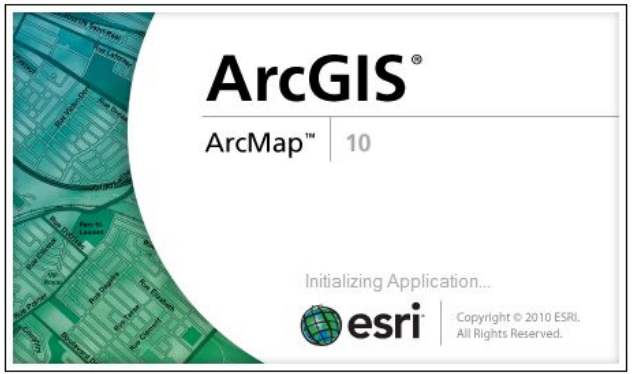




GIS software and analytical tools

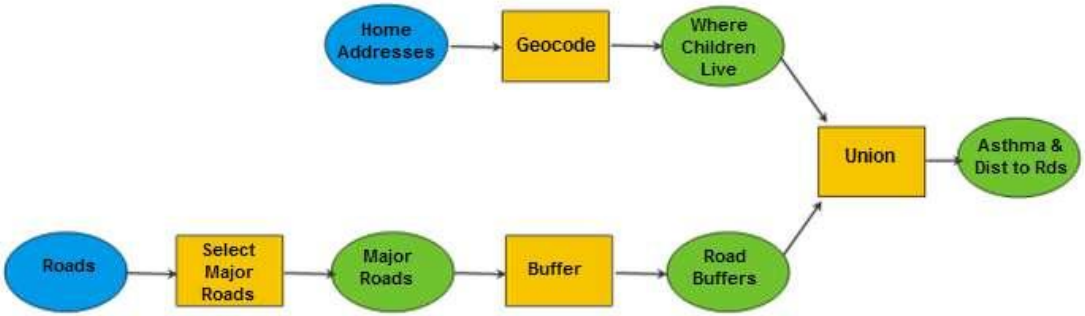
GRASS:
Open source GIS

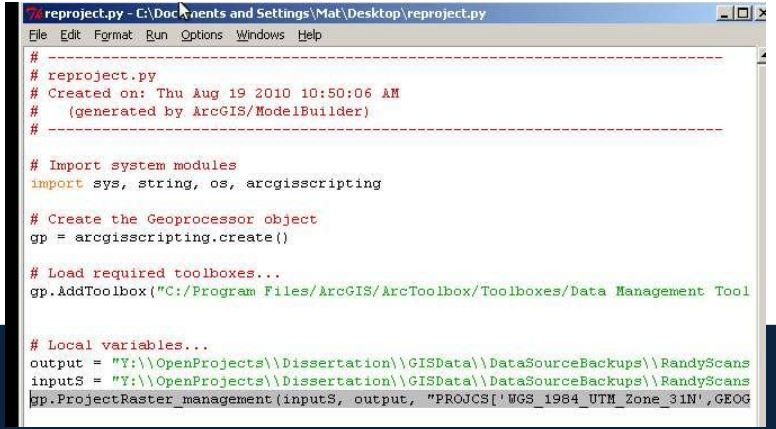
Quantum GIS:
Open source GIS



ArcInfo GIS:
Proprietary, industry standard, GIS package

Model building and Scripting





https://github.com/msisk1/workshops/tree/master/SpatialR