Introduction to GIS and Spatial Analysis

July 13, 2021

Parmanand Sinha

Computational Scientist, Research Computing Center

About RCC-GIS

- **Support for Geospatial Information Science Research**
- Assistance for all UChicago students, faculty, and staff
  - Training
  - Consultation
 - Events

- https://gis.rcc.uchicago.edu

Topics

- Software and packages for GIS analysis
- Basic GIS operations.
- Understanding the Census
- Spatial Analysis with Vector data
- Geocoding and Geo-referencing
- Basic web mapping techniques

GIS

- A geographic information system (GIS) lets us visualize, question, analyze, interpret, and **understand data to reveal relationships, patterns, and trends**.

  - - esri

Click to add text
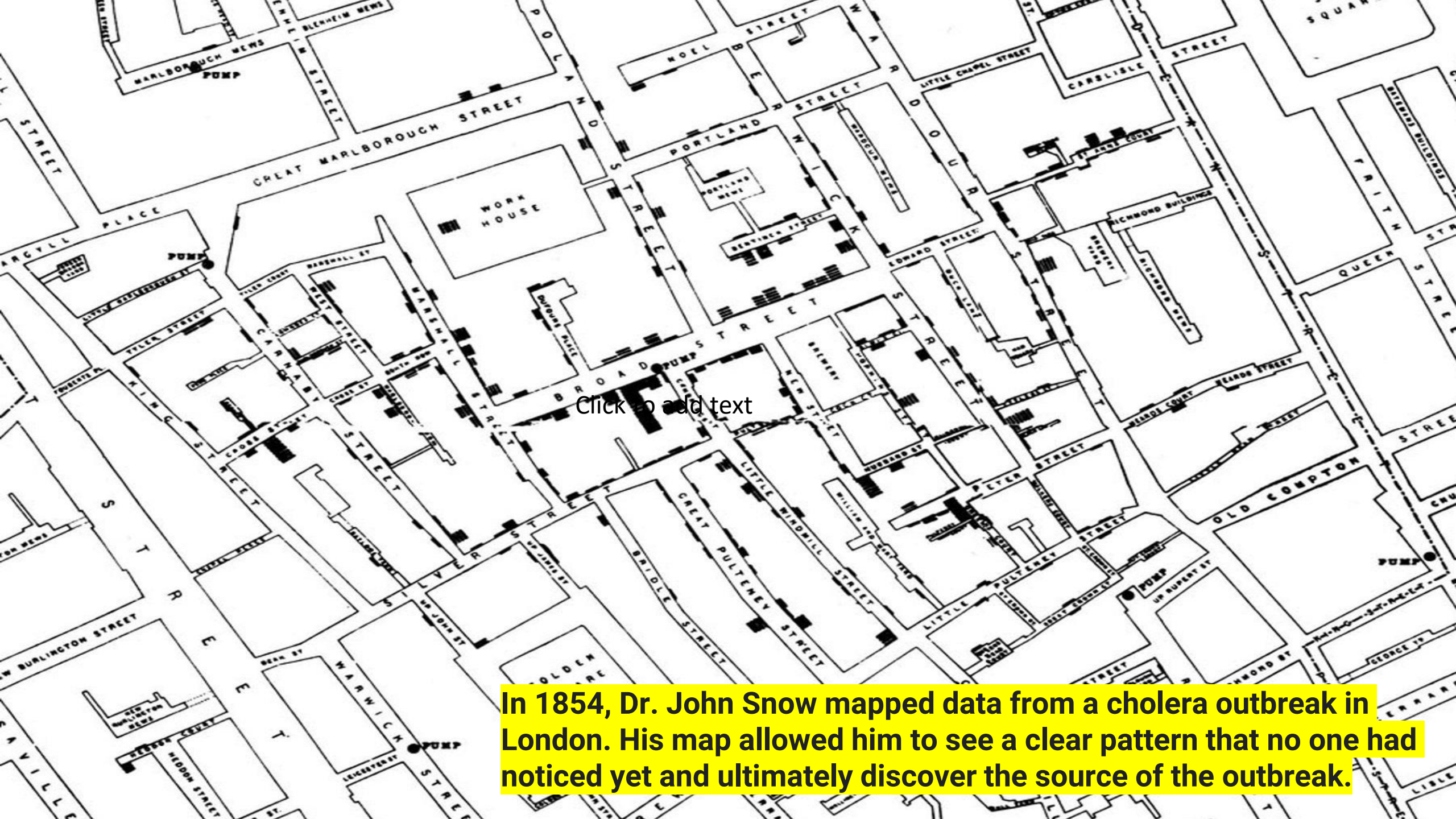
In 1854, Dr. John Snow mapped data from a cholera outbreak in London. His map allowed him to see a clear pattern that no one had noticed yet and ultimately discover the source of the outbreak.

Maps + Data = GIS

As an integrated system for geographic data

- Capturing data
- Storing data
- Querying data
- Analyzing data
- Displaying data

What can GIS do?
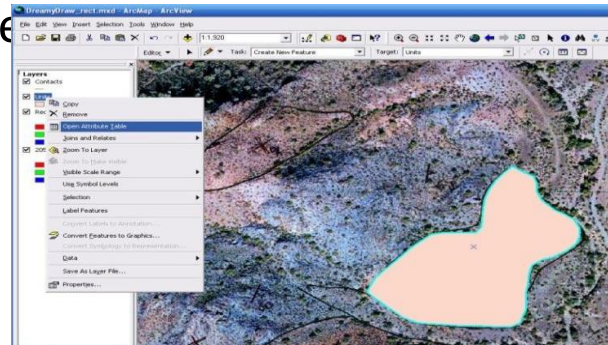
What can GIS do?

- Geographic Information System (GIS):

  An integrated system for geographic data

    - Capturing data
    - Storing data
    - Querying data
    - Analyzing data
    - Displaying data

Capturing data

- Digitizing

  - Creating a digital copy of existing data

  - Paper maps, Aerial imagery, Topographic maps

  - Output is a GIS friendly vector format

- GPS

  - Global Positioning System

  - Coverage of entire planet

  - Data Formats:
    - Temporal – accurate to about 14 nanoseconds
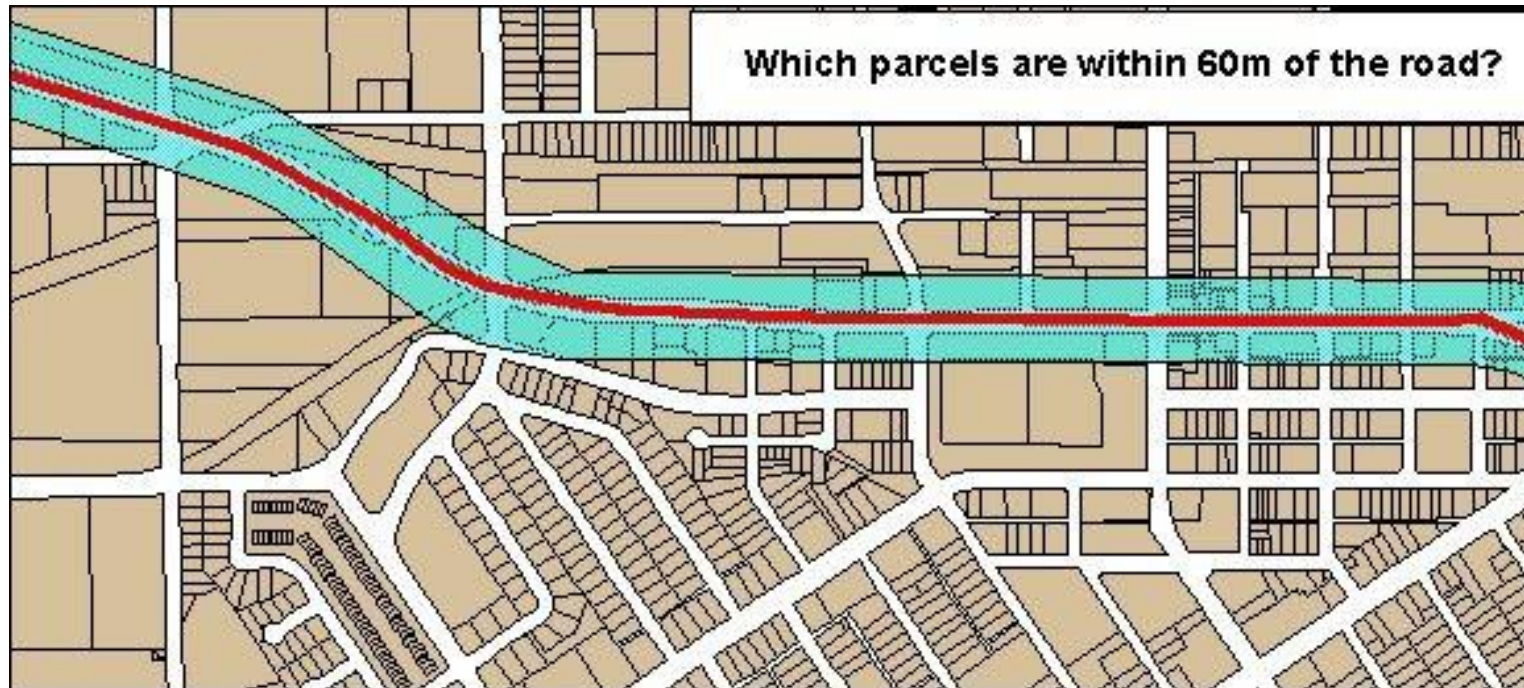   - Spatial – sub-meter accuracy

Storing data

- Files
  - Shapefile
  - KML (Google)
  - Spreadsheet
  - Geodatabase
- Spatial Database
  - PostGIS, Oracle, SQL Server
  - Traditional database management system that also stores vector feature geometry and location data

Querying data

Ask a question of the data based on location or other attribute

Which parcels are within 60m of the road?

Analyzing data

- Performing operations on the spatial data
  - "How far am I from the nearest park?"
  - Overlays
  - Extracts
  - Proximity
  - Spatial Statistics

Obesity rates in school zones

Obesity density surface

Predicted rates for census blocks
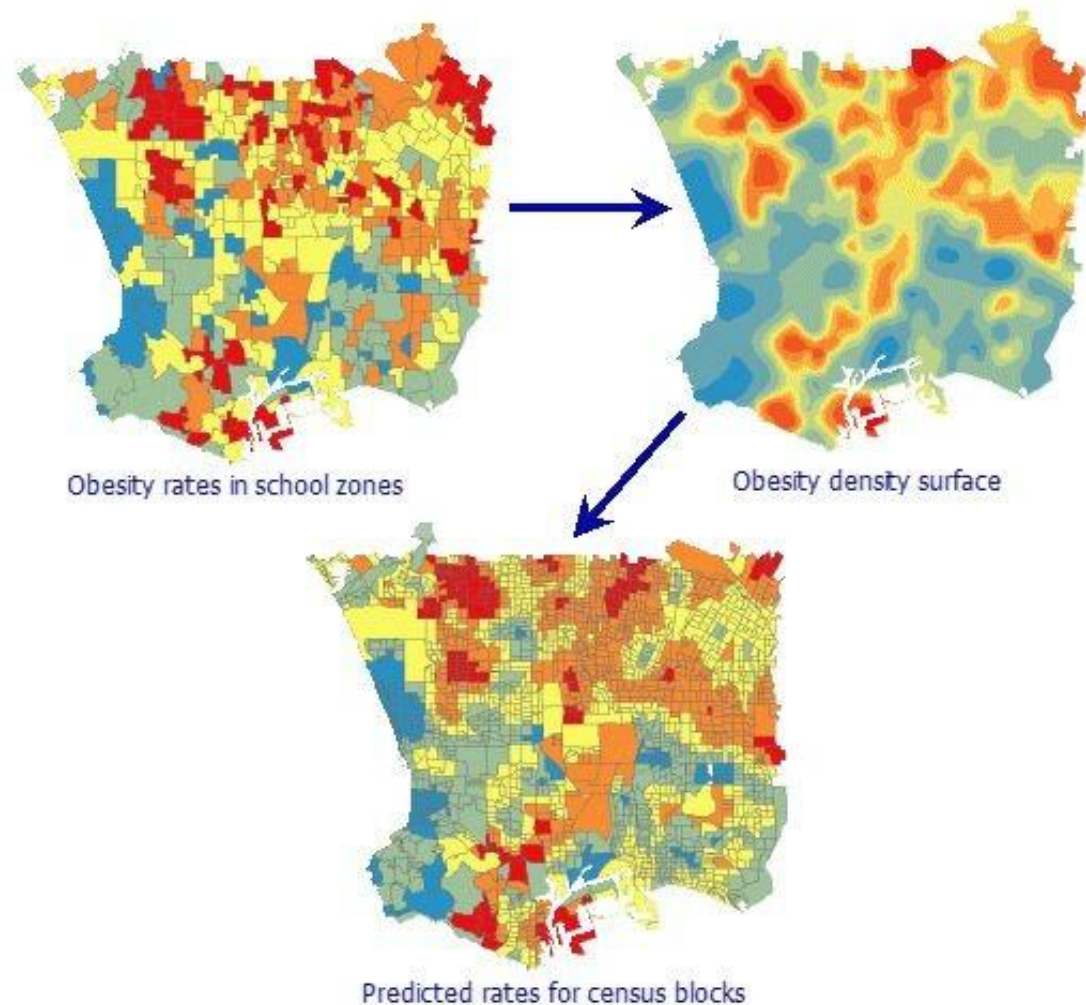
Displaying data

- Mapping
  - The display of spatial data for a specific purpose
  - Visualization
    - Re-emerging as a popular means to artistically portray traditional data
    - Not necessarily spatial data
- Output Format
  - Print or Web
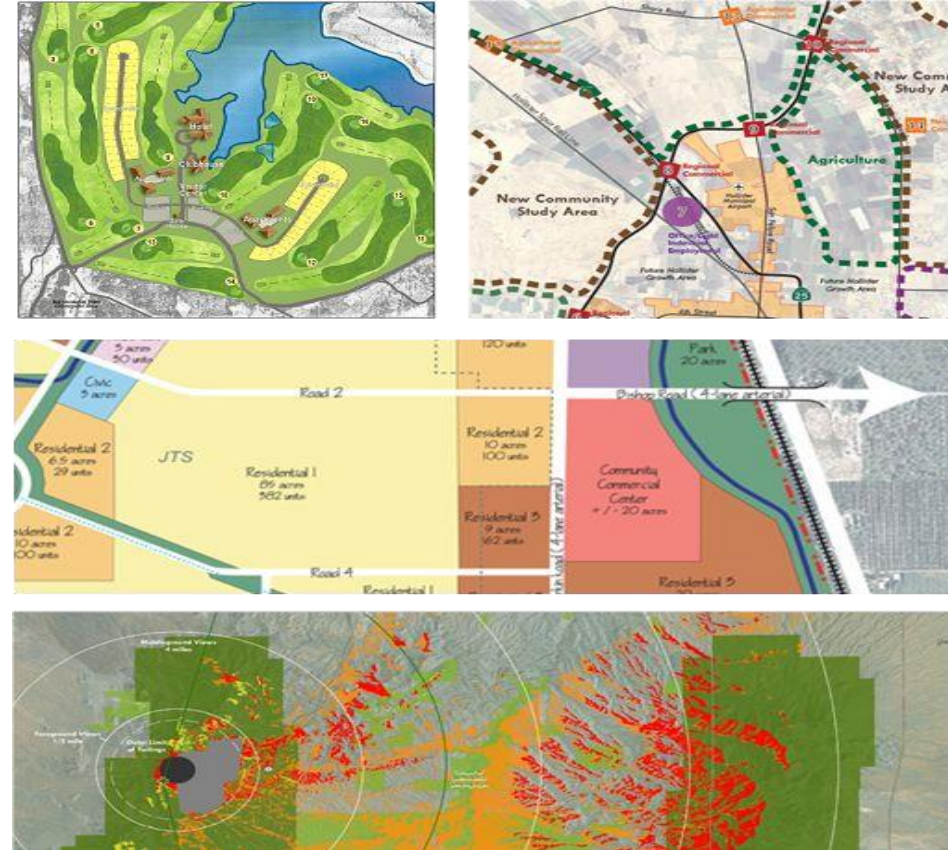
Spatial Data Models

- Vector

- Raster
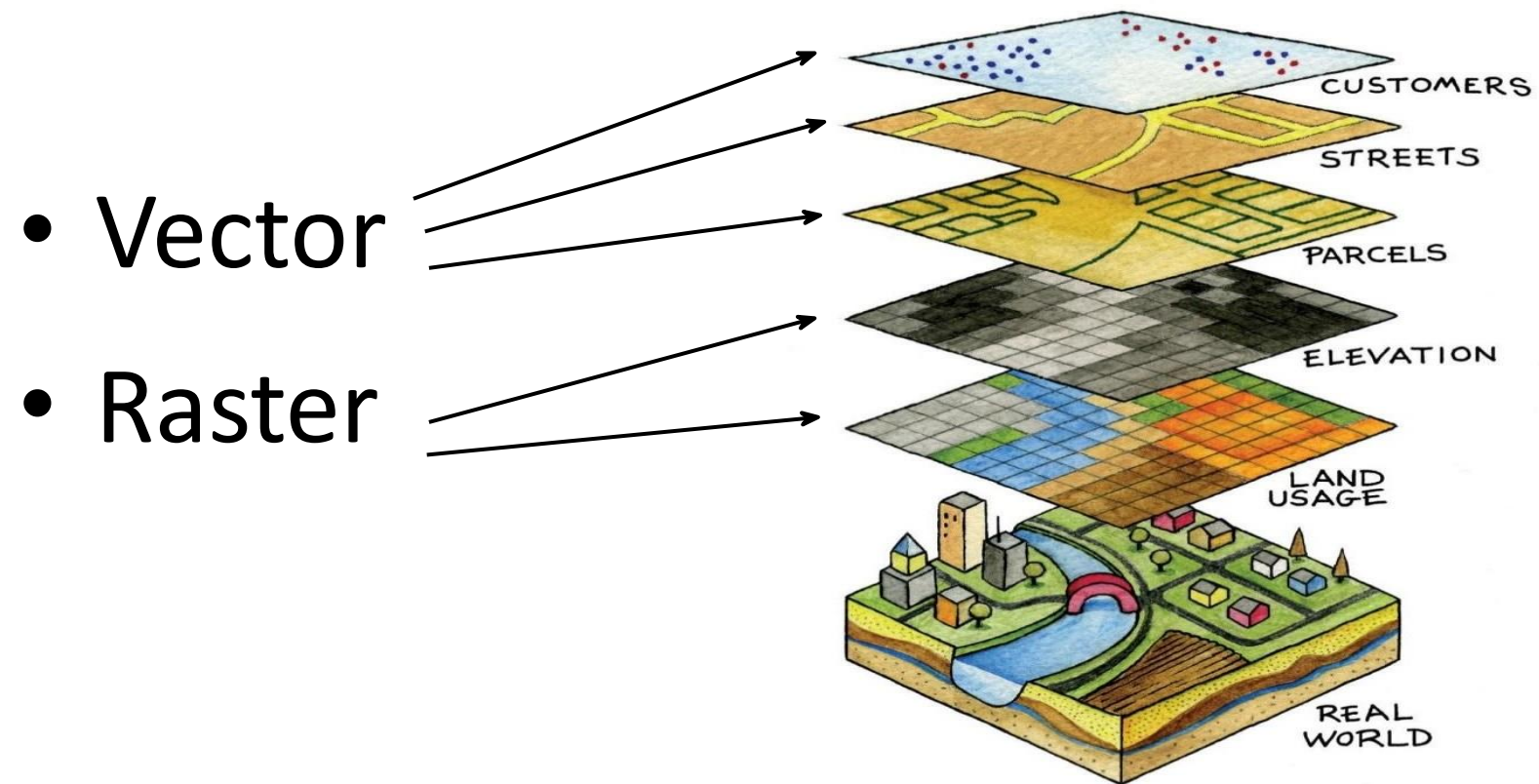
Vector Data Model
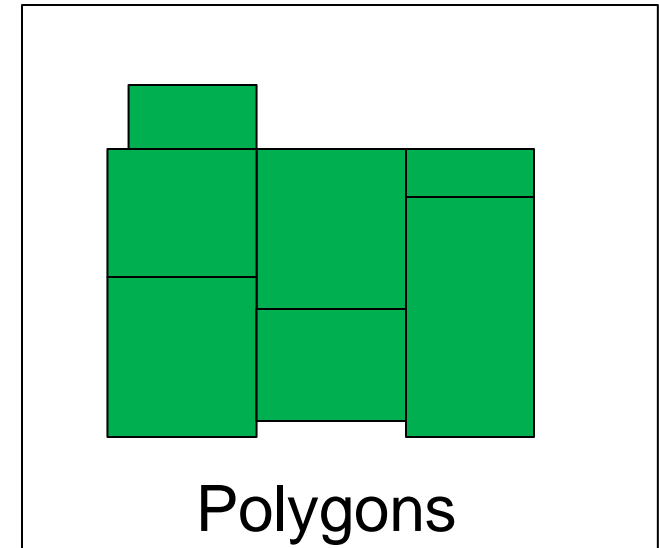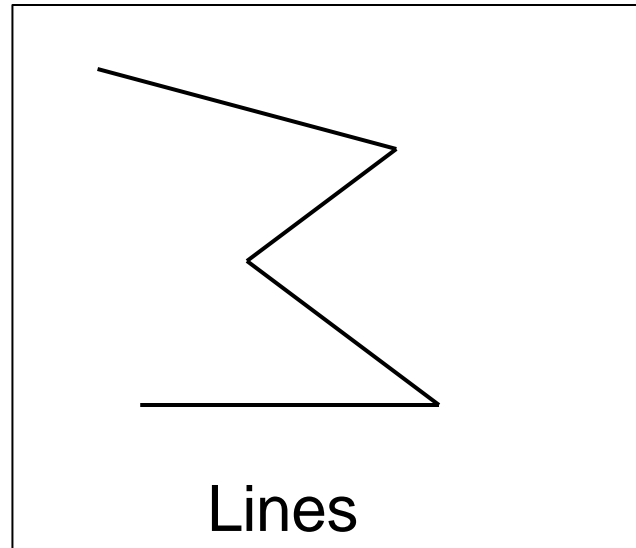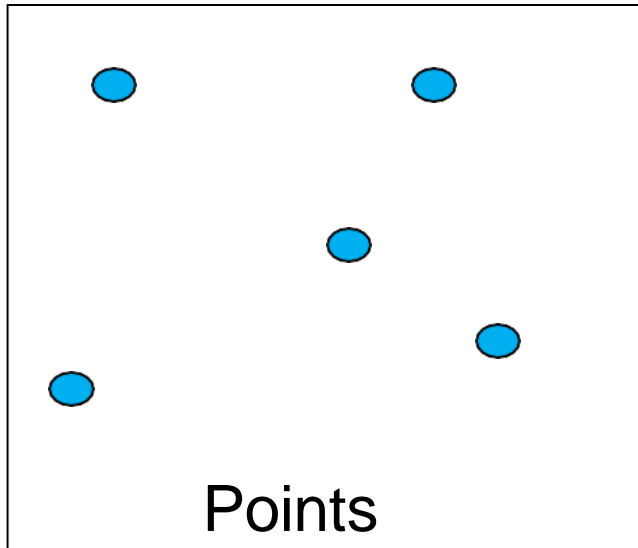
Spatial and Tabular Data

- Adds geometry to traditional data
  - Geometry: location and geometric characteristics of geographic (real-world) features
  - Attributes: data describing the characteristics of geographic features

Spatial and Tabular Data

Geometry + Attributes
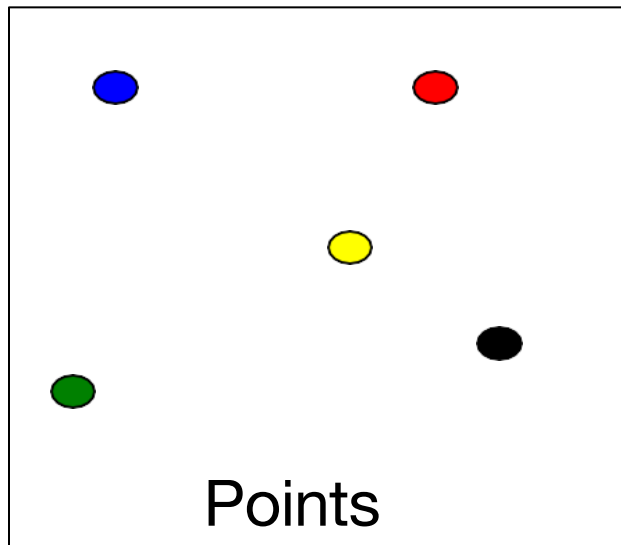
Points

ID	Color	Use
1	Blue	Hospital
2	Red	Fire Dept.
3	Green	Office
4	Yellow	University
5	Black	Retail

Vector Model

- Vector
  - Discrete entities defined by coordinate points
  - Three types of vector data
    - Point
    - Line
    - Polygon

18
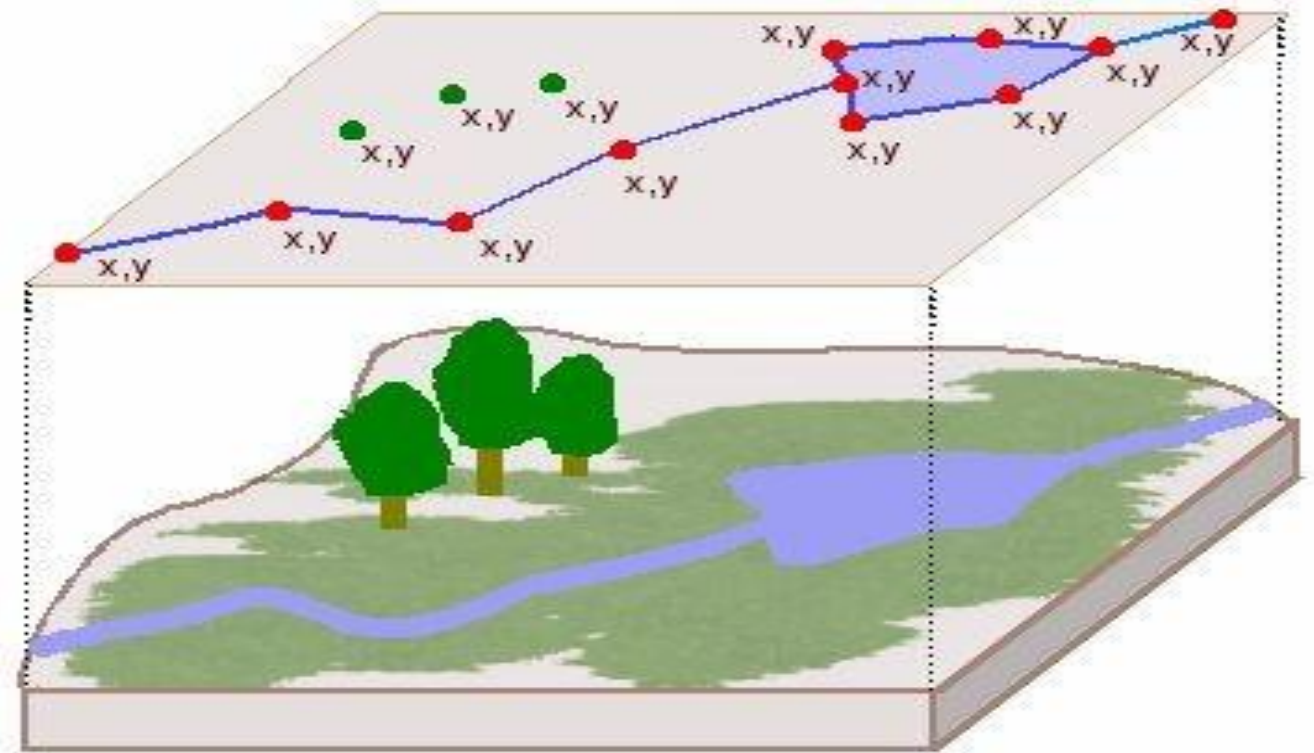
Raster Model

- Composed of a regular grid of cells
- Every grid cell has a value
- Every point on ground belongs to a grid cell
- Examples
  - Elevation
  - Crime hotspot
  - Temperature
  - Rainfall

AVERAGE ELEVATION
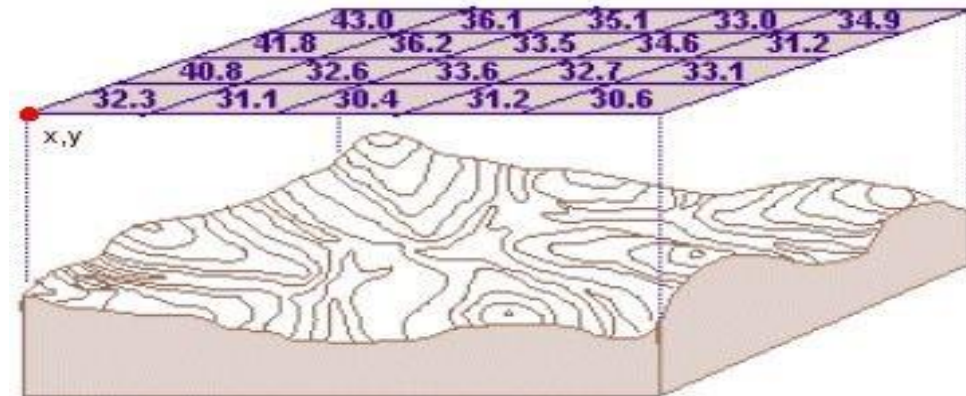
Raster Data Model

- Conventionally, stored row by row from the top left corner

- Attributes are recorded by assigning each cell a single value: e.g., landuse type

- Simple data structure

  - Directly store each layer as a single table

  - each layer is analogous to a ''spreadsheet'' or ''matrix''
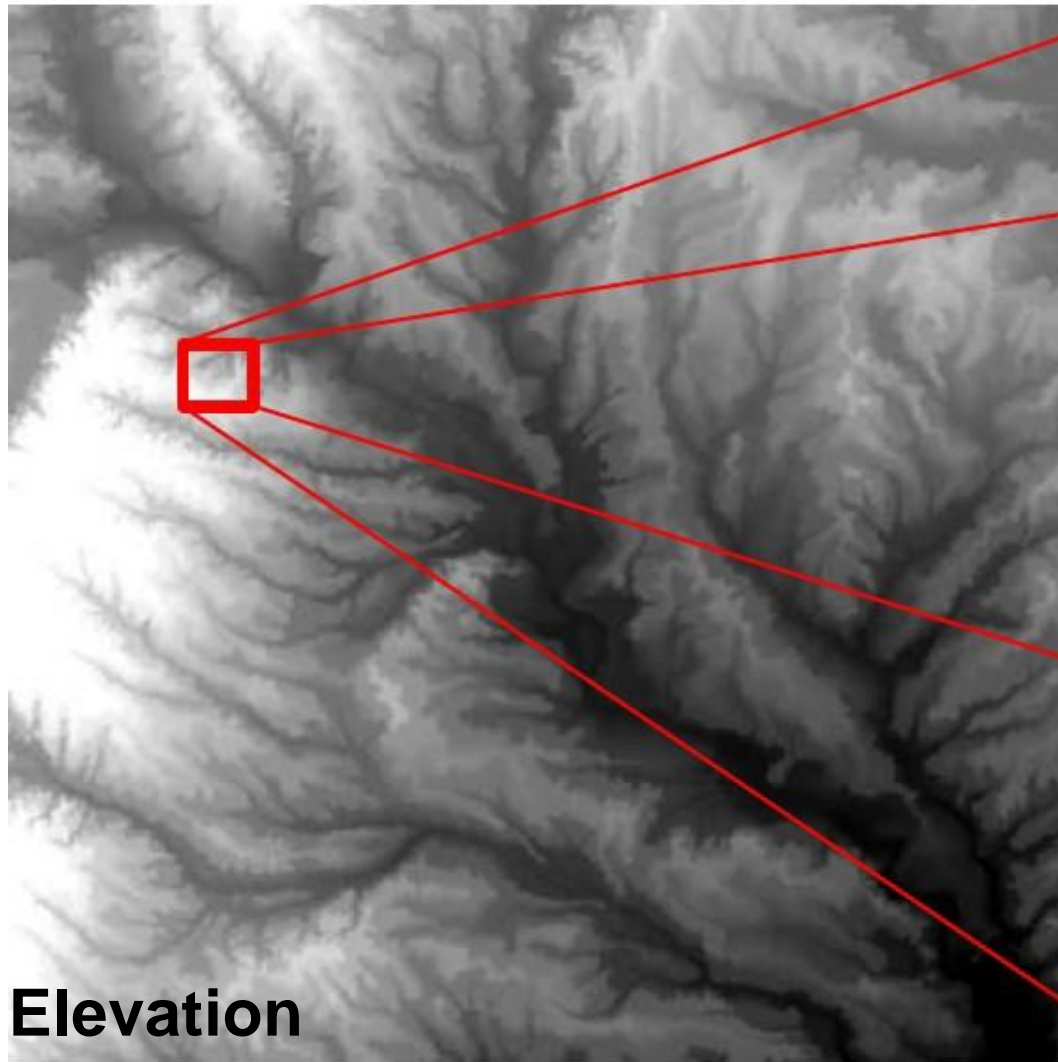
Elevation

- High : 262

- Low: 73

Elevation in Dallas county

The whole county and a small area in thecounty
(Data Source: USGS)

Raster vs. Vector

- ## Different ways of displaying the "real world"

Raster / Image

Vector

Real World
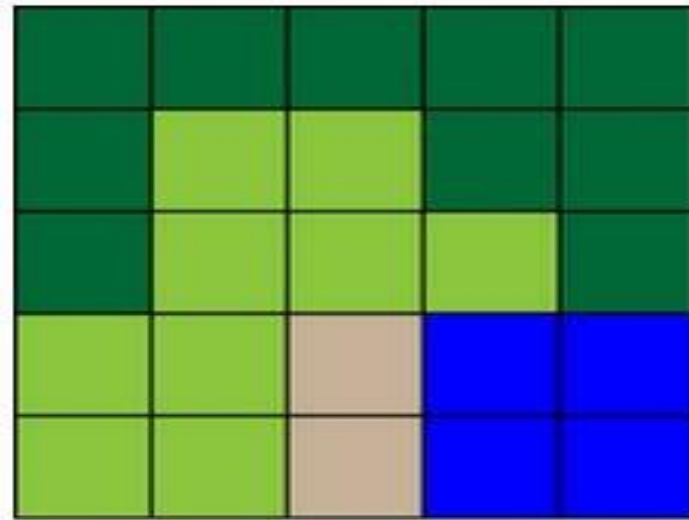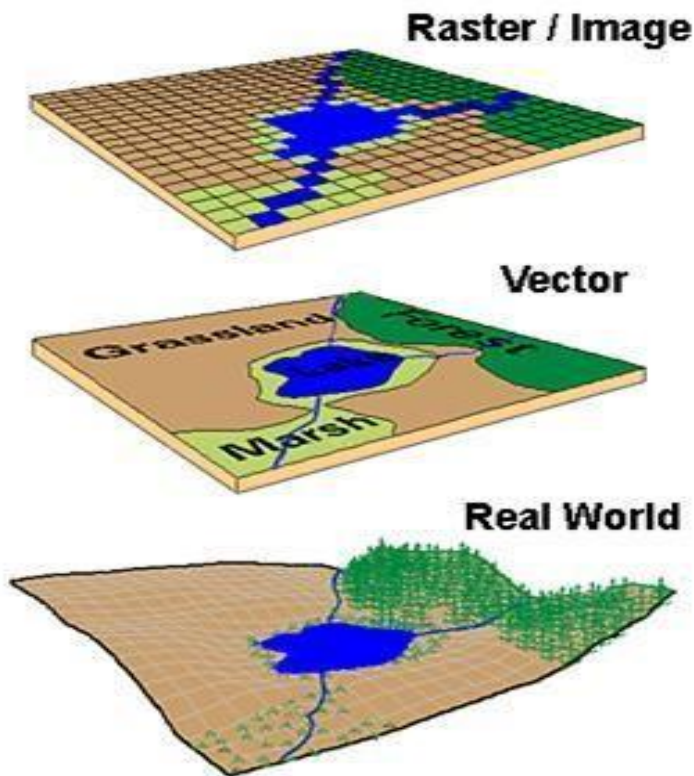
Raster

Vector

Generalization

- The amount of detail given to an object

Raster   Vector
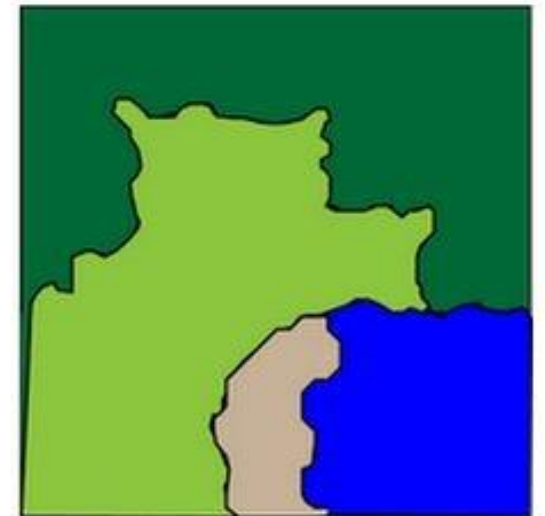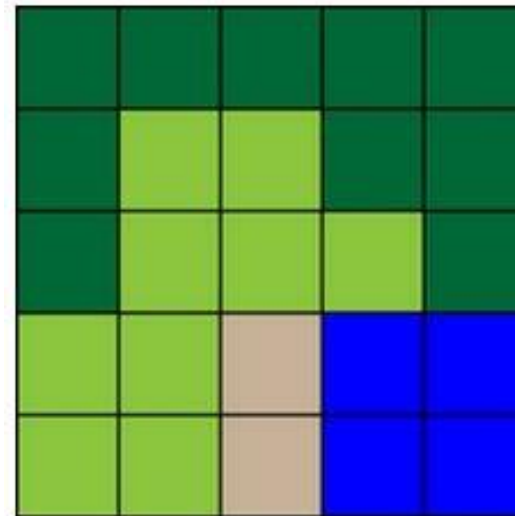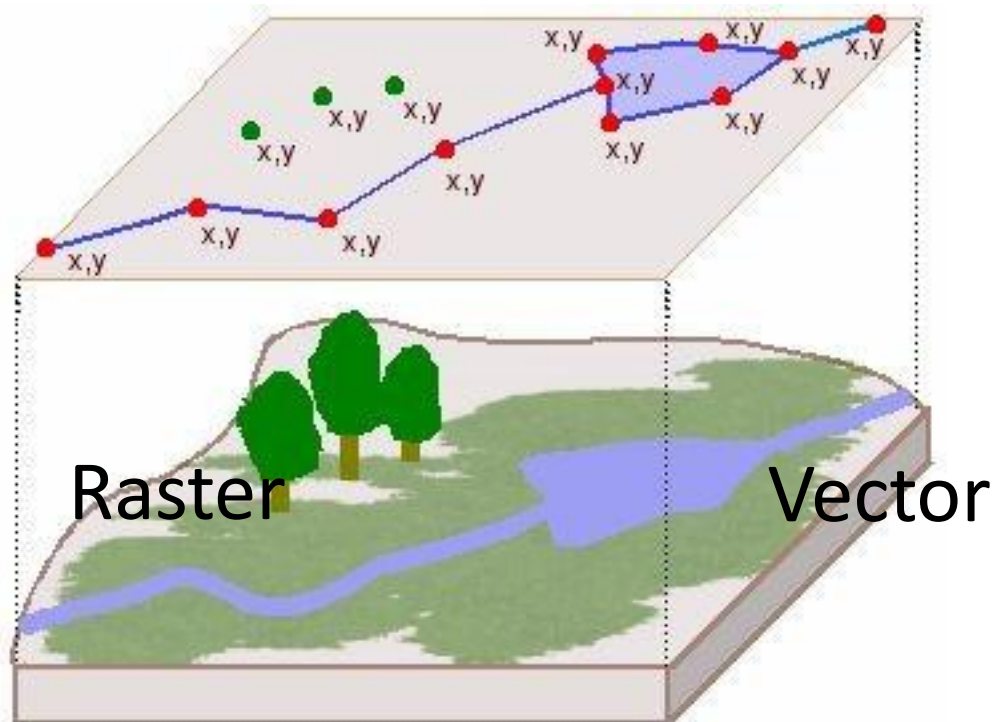
Data formats

Individual files vs. databases

Spatial Data: Shapefile

- Most common spatial data format in GIS

- Has been around for 40 years!

- Made up of points, lines, or polygons (**vector**)

- All GIS software will read shapefiles

- Used across all disciplines

Spatial Data: Shapefile

- Single shapefile actually consists of multiple files
  - <u>.shp</u> – stores geometry
  - <u>.dbf</u> – stores attributes
  - <u>.shx</u> – index file
  - .prj – projection file
  - .xml – metadata file

bike_trails.dbf	10/8/2017 2:47 PM	DBF File	164 KB
bike_trails.prj	10/8/2017 2:47 PM	PRJ File	1 KB
bike_trails.qpj	10/8/2017 2:47 PM	QPJ File	1 KB
bike_trails.shp	10/8/2017 2:47 PM	SHP File	300 KB
bike_trails.shx	10/8/2017 2:47 PM	SHX File	5 KB

Geodatabase

- Stores a set of files

- Also allows for data query, data management

Canada
- Airports
- Cities
- Elev
- Lakes
- Parks
- PopStats
- Provinces
- Railroads
- Roads

Geodatabase with feature classes, a raster dataset, and a stand-alone table

Feature Class

- Layer
- Grouping of one type of feature (i.e. points, lines, polygons)

- With spatial and attribute information for each feature
- A "shapefile" stored within a Geodatabase

Metadata

- Data about data
- Describes the content, lineage, creator, distributor, processing steps, and spatial reference of the spatial data
- Helps users determine the availability and access requirements for data
- Helps users judge the quality and "fitness-for-use" of the data for their particular application
- Results only as good as input data!
- "**View Metadata**" in ArcGIS Pro

Projection

- Earth is spherical, maps are not

- Projections allow for "best" views of specific areas

- Will not be covered in detail but

  - Different Scales use different projections

  - States have their own projections

  - Counties have their own projections

Projections

projection distortion

- The Earth is an ellipsoid whose larger radius is located at the Equator. If the jurisdiction of the owner is small enough, the planar assumption is valid.

No distortion

Map

Earth surface

Most distortion

Most distortion

The **True Size** of Africa
A small contribution in the fight against rampant *Immappancy,* by **Kai Krause**

In addition to the well known social issues of *illiteracy* and *innumeracy,* there also should be such a concept as *"immappancy",* meaning *insufficient geographical knowledge.*

A survey with random American schoolkids let them guess the population and land area of their country. Not entirely unexpected, but still rather unsettling, the majority chose *"1-2 billion"* and *"largest in the world",* respectively. Even with Asian and European college students, geographical estimates were often off by factors of *2-3*. This is partly due to the highly distorted nature of the predominantly used mapping projections (such as *Mercator*).

A particularly extreme example is the worldwide misjudgement of the true size of Africa. This single image tries to embody the massive scale, which is larger than the *USA, China, India, Japan* and all of Europe - combined!

COUNTRY	AREA x 1000 km²
USA	9.629
China	9.573
India	3.287
Mexico	1.964
Peru	1.285
France	633
Spain	506
Papua New Guinea	462
Sweden	441
Japan	378
Germany	357
Norway	324
Italy	301
New Zealand	270
United Kingdom	243
Nepal	147
Bangladesh	144
Greece	132
TOTAL	30.102
AFRICA	30.221
Just for Reference: The Surface of the MOON	37.930

Please note:

The graphical layout of this map is meant purely as a *visualization* to illustrate the fact: Africa is *much* larger than *almost everyone* assumes!

Even totally blurred outlines could have been used to make that point, however the table at left is very accurate, citing:

http://en.wikipedia.org/wiki/List_of_countries_and_outlying_territories_by_total_area

Note for instance that the figure in the table for the USA *does* include Alaska and Hawaii, but they *are not even used* in the map, as are a handful of other entries (such as Norway and Sweden).

The reason for this is that the map purposely uses the familiar shapes, as if you are 'moving pieces in Google Maps'. Because the mathematically exact depiction, using equal area scaling, would be *even more drastic,* but would appear highly distorted. I chose to retain the commonly known outlines and proportions to tell the story, even if this conservative size has 'left-over parts'.

The small maps on the right are again the singular message: see some of the countries in direct relation to Africa, a view that is quite unfamiliar and rarely seen.

It is worth looking at *Bucky Fullers* maps or the *Peters* equal area proposals, among many other beautiful attempts to display geographical information. Numerous other side-by-side comparisons have been made, this is by far not the first and hopefully *not the last* such map: someone should find the best fit of all puzzle pieces in a neutral projection.

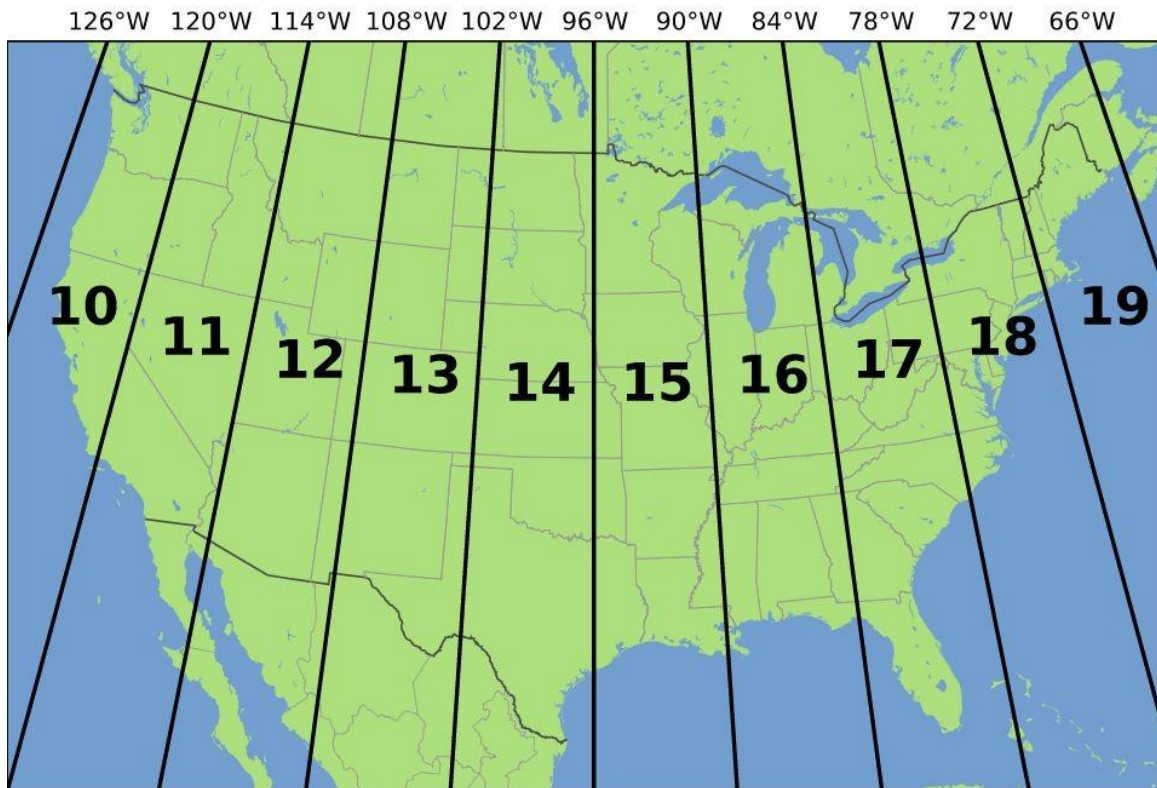Until then, please do not take it all *too literal* ('where is Ibiza??') and simply take that one impression with you: Africa.... is immense.

Top 100 Countries
Area in square kilometers, Percentage of World Total
Sources: Britannica, Wikipedia, Almanac 2010

		AREA km²	%
1	Russia	17.098.242	11,50
2	Canada	9.984.670	6,70
3	China	9.596.961	6,40
4	United States	9.629.091	6,40
5	Brazil	8.514.877	5,70
6	Australia	7.692.024	5,20
7	India	3.287.263	2,30
8	Argentina	2.780.400	2,00
9	Kazakhstan	2.724.900	1,80
10	Sudan	2.505.813	1,70
11	Algeria	2.381.741	1,60
12	Congo	2.344.858	1,60
13	Greenland	2.166.086	1,50
14	Saudi Arabia	2.149.690	1,40
15	Mexico	1.964.375	1,30
16	Indonesia	1.860.360	1,30
17	Libya	1.759.540	1,20
18	Iran	1.628.750	1,10
19	Mongolia	1.564.100	1,10
20	Peru	1.285.216	0,86
21	Chad	1.284.000	0,86
22	Niger	1.267.000	0,85
23	Angola	1.246.700	0,85
24	Mali	1.240.192	0,83
25	South Africa	1.221.037	0,82
26	Colombia	1.141.748	0,76
27	Ethiopia	1.104.300	0,74
28	Bolivia	1.098.581	0,74
29	Mauritania	1.025.520	0,69
30	Egypt	1.002.000	0,67
31	Tanzania	945.087	0,63
32	Nigeria	923.768	0,62
33	Venezuela	912.050	0,61
34	Namibia	824.116	0,55
35	Mozambique	801.590	0,54
36	Pakistan	796.095	0,53
37	Turkey	783.562	0,53
38	Chile	756.102	0,51
39	Zambia	752.612	0,51
40	Myanmar	676.578	0,45
41	Afghanistan	652.090	0,44
42	Somalia	637.657	0,43
43	France	632.834	0,43
44	C. African Rep	622.984	0,42
45	Ukraine	603.500	0,41
46	Madagascar	587.041	0,39
47	Botswana	582.000	0,39
48	Kenya	580.367	0,39
49	Yemen	527.968	0,35
50	Thailand	513.120	0,34
51	Spain	505.992	0,34
52	Turkmenistan	488.100	0,33
53	Cameroon	475.442	0,32
54	Papua New Guinea	462.840	0,31
55	Uzbekistan	447.400	0,30
56	Morocco	446.550	0,30
57	Sweden	441.370	0,30
58	Iraq	438.317	0,29
59	Paraguay	406.752	0,27
60	Zimbabwe	390.757	0,26
61	Japan	377.930	0,25
62	Germany	357.114	0,24
63	Rep o.t. Congo	342.000	0,23
64	Finland	338.419	0,23
65	Vietnam	331.212	0,22
66	Malaysia	330.803	0,22
67	Norway	323.802	0,22
68	Côte d'Ivoire	322.463	0,22
69	Poland	312.685	0,21
70	Oman	309.500	0,21
71	Italy	301.336	0,20
72	Philippines	300.000	0,20
73	Burkina Faso	274.222	0,18
74	New Zealand	270.467	0,18
75	Gabon	267.668	0,18
76	Western Sahara	266.000	0,18
77	Ecuador	256.369	0,20
78	Guinea	245.857	0,17
79	United Kingdom	242.900	0,16
80	Uganda	241.038	0,16
81	Ghana	238.539	0,16
82	Romania	238.391	0,16
83	Laos	236.800	0,16
84	Guyana	214.969	0,14
85	Belarus	207.600	0,14
86	Kyrgyzstan	199.951	0,13
87	Senegal	196.722	0,13
88	Syria	185.180	0,12
89	Cambodia	181.035	0,12
90	Uruguay	176.215	0,12
91	Suriname	163.820	0,11
92	Tunisia	163.610	0,11
93	Nepal	147.181	0,10
94	Bangladesh	143.998	0,10
95	Tajikistan	143.100	0,10
96	Greece	131.957	0,09
97	Nicaragua	130.373	0,09
98	North Korea	120.538	0,08
99	Malawi	118.484	0,08
100	Eritrea	117.600	0,08
	TOP 100 TOTAL	133.882.924	89,34

United States

Europe

India

Japan

China

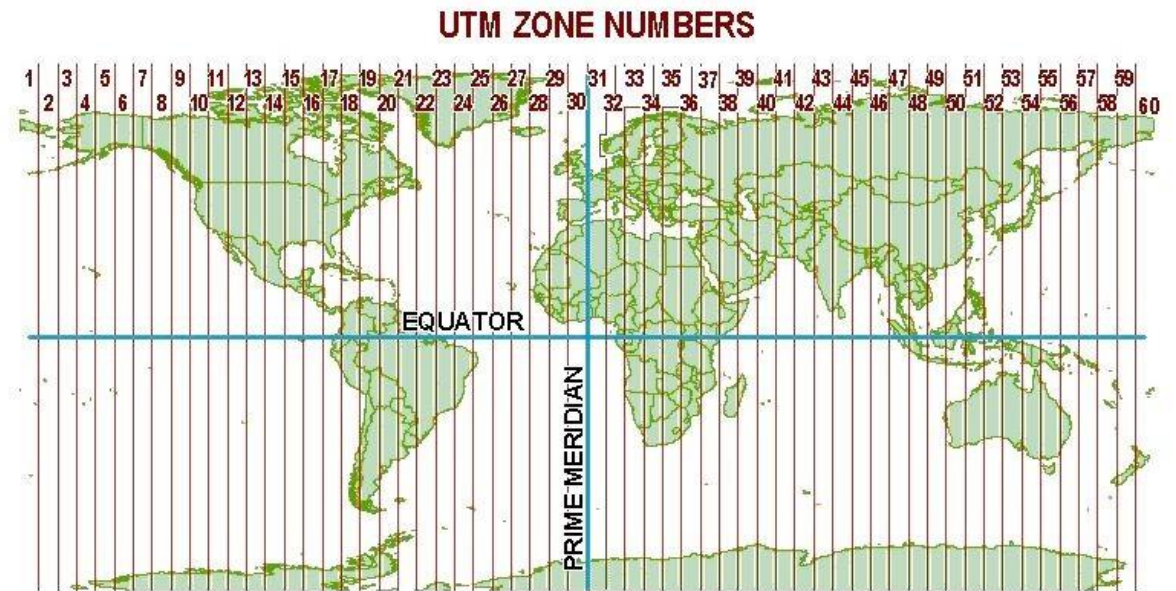Different projections: The Mercator projection is very biased, as you can see from The True Size of Africa.

UTM Projections

UTM Zones for the United States

UTM Zones for the World

Mapping

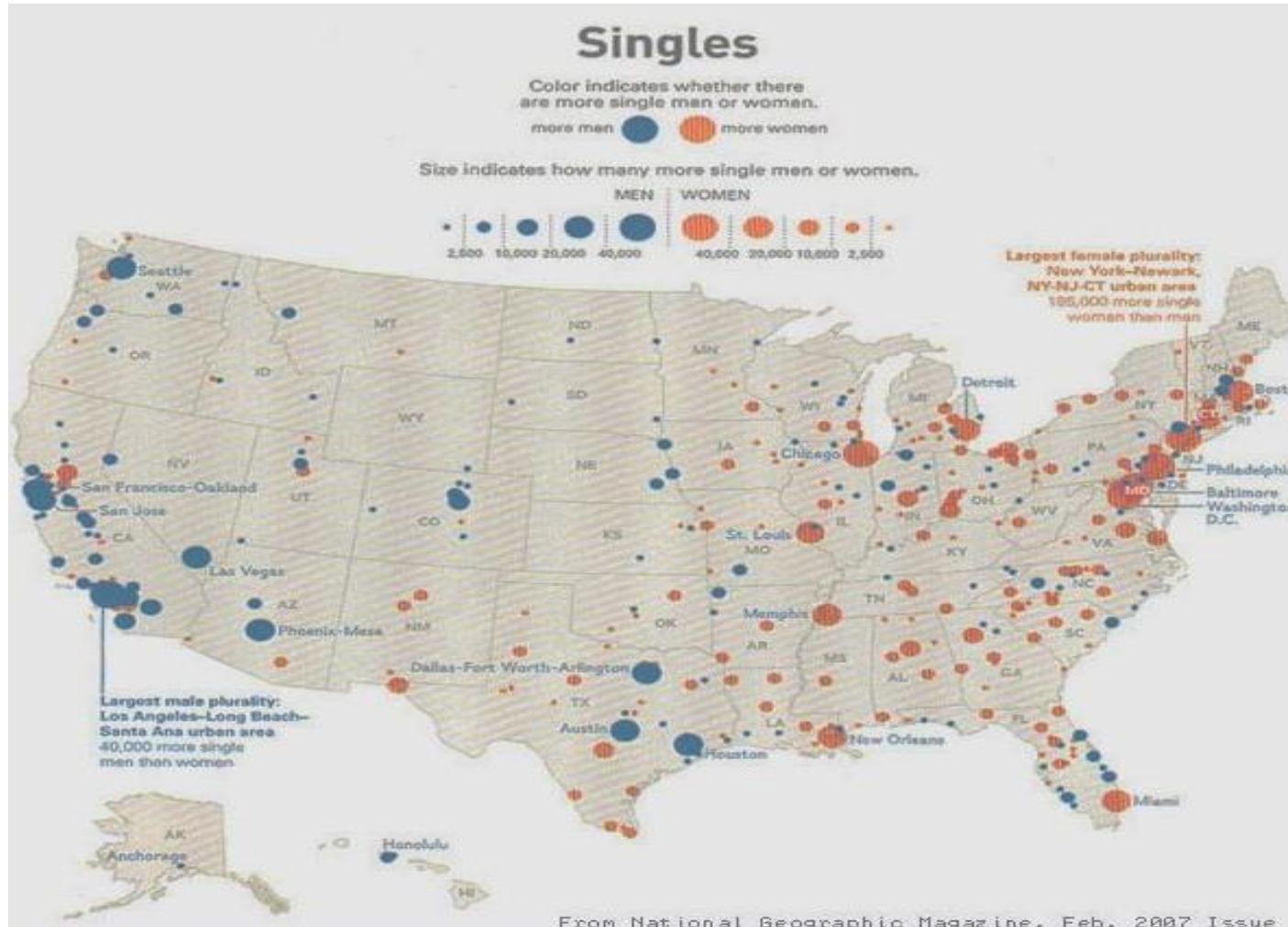# Types of Thematic Maps

- Choropleth

- Graduated Symbol

- Hot Spot

Choropleth Map

Percent of Farms with Sales
of $250,000 or More: 2002

Percent
Less than 1
1 – 4
5 - 9
10 - 14
15 or more

United States
7.2 Percent

02-M011
U.S.Department of Agriculture, National Agricultural Statistics Service

Shows value per unit using colors. Use these for comparing relative numbers across space.

Proportional Symbol Map

From National Geographic Magazine, Feb. 2007 Issue

The size of the symbol is representative of the value of the variable.

Difficult use well.

Hot Spot Map

Shows density or clustering of phenomenon using color.

Statistically significant areas are displayed, excluding areas with no data.

Data Standardization

- Data should usually be standardized

  - Ability to compare areas

- By Population

  - Per person, Percentages

- By Area

  - Density

Data Standardization by Pop

Unstandardized

Standardized/Normalized

Korean War Veterans, 1970

- 46 - 625
- 626 - 2053
- 2054 - 5114
- 5115 - 10686
- 10687 - 24288

Korean War Veterans, 1970

- 0.16% - 1.54%
- 1.55% - 2.16%
- 2.17% - 2.7%
- 2.71% - 4.44%
- 4.45% - 14.42%
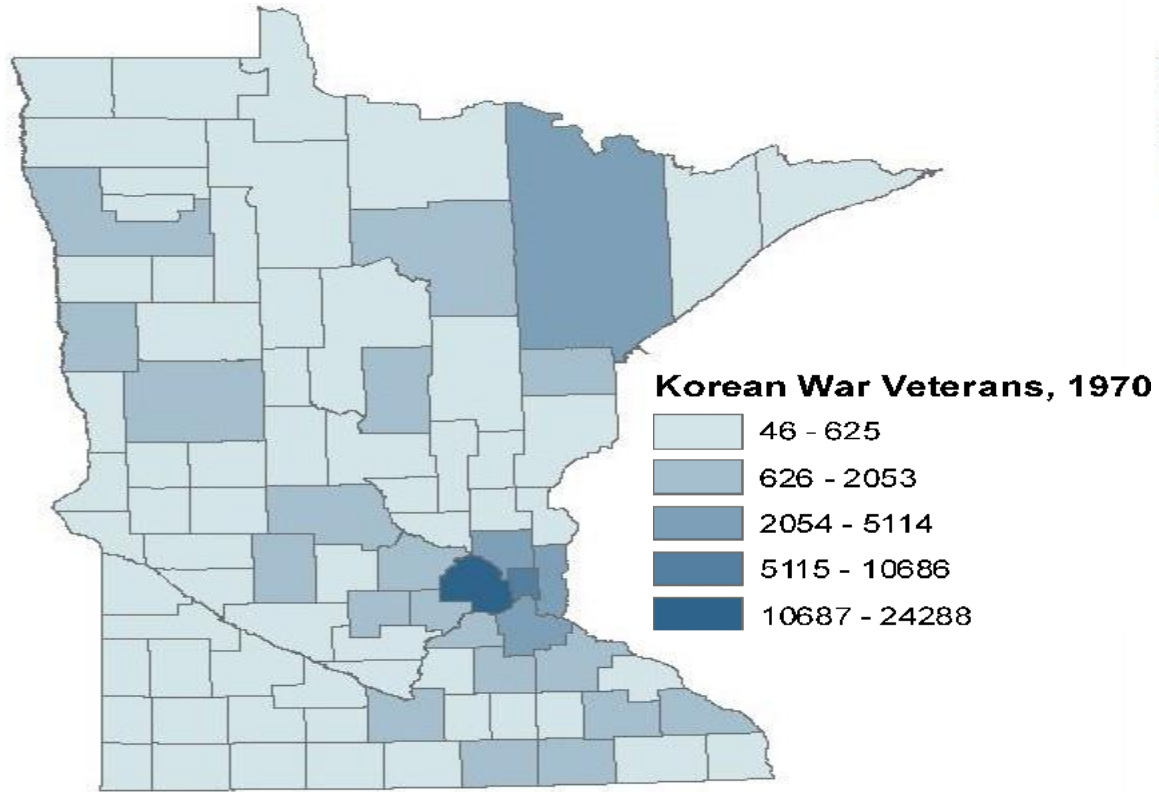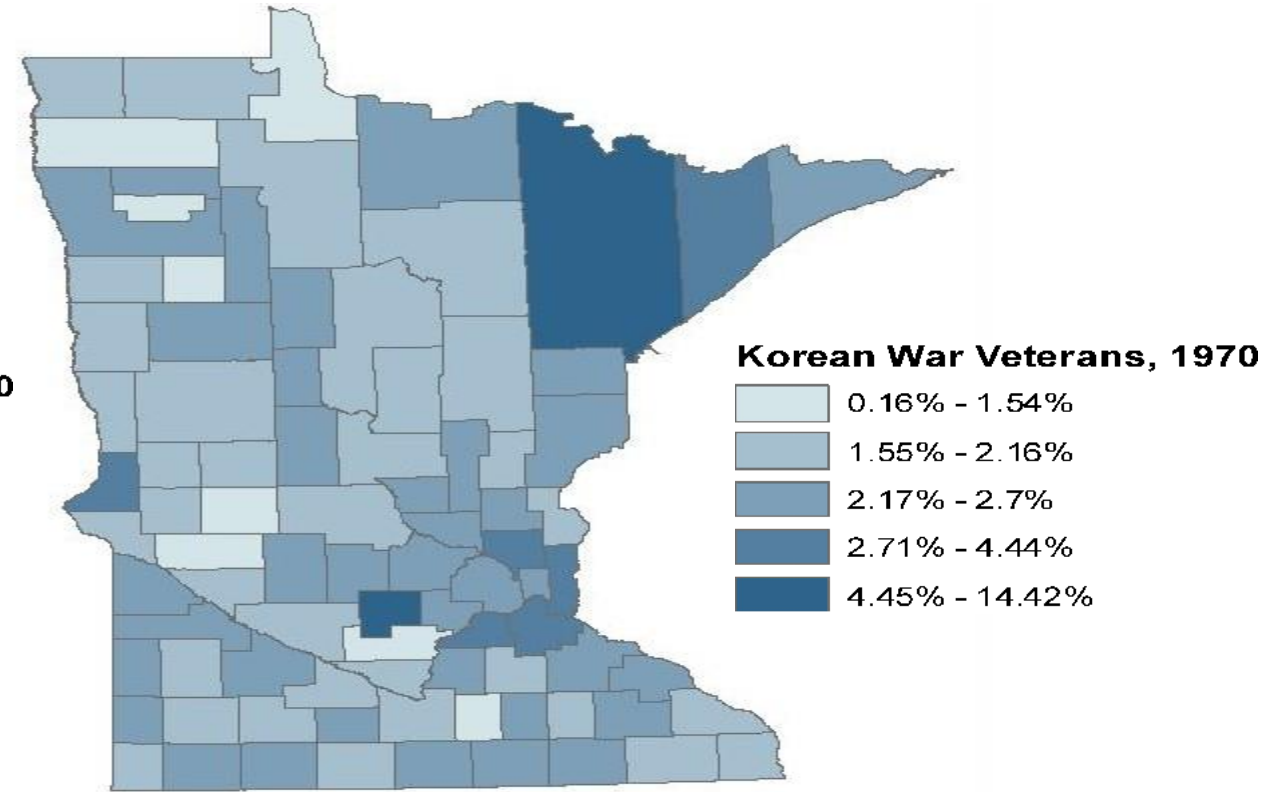
Data Standardization by Area

Unstandardized

Standardized/Normalized

Population in 1970
- 3423 - 23208
- 23209 - 52322
- 52323 - 220693
- 220694 - 476255
- 476256 - 960080

Population in 1970, per SqMi
- 2.131 - 25.70
- 25.71 - 54.50
- 54.51 - 128.5
- 128.6 - 367.6
- 367.7 - 2799
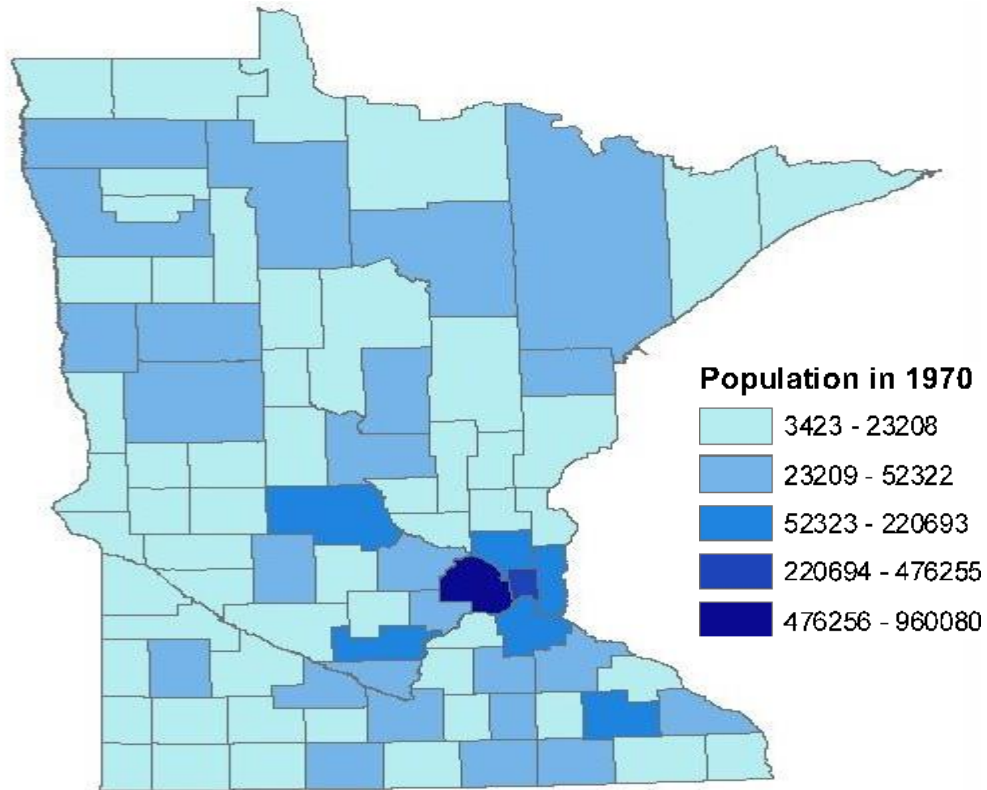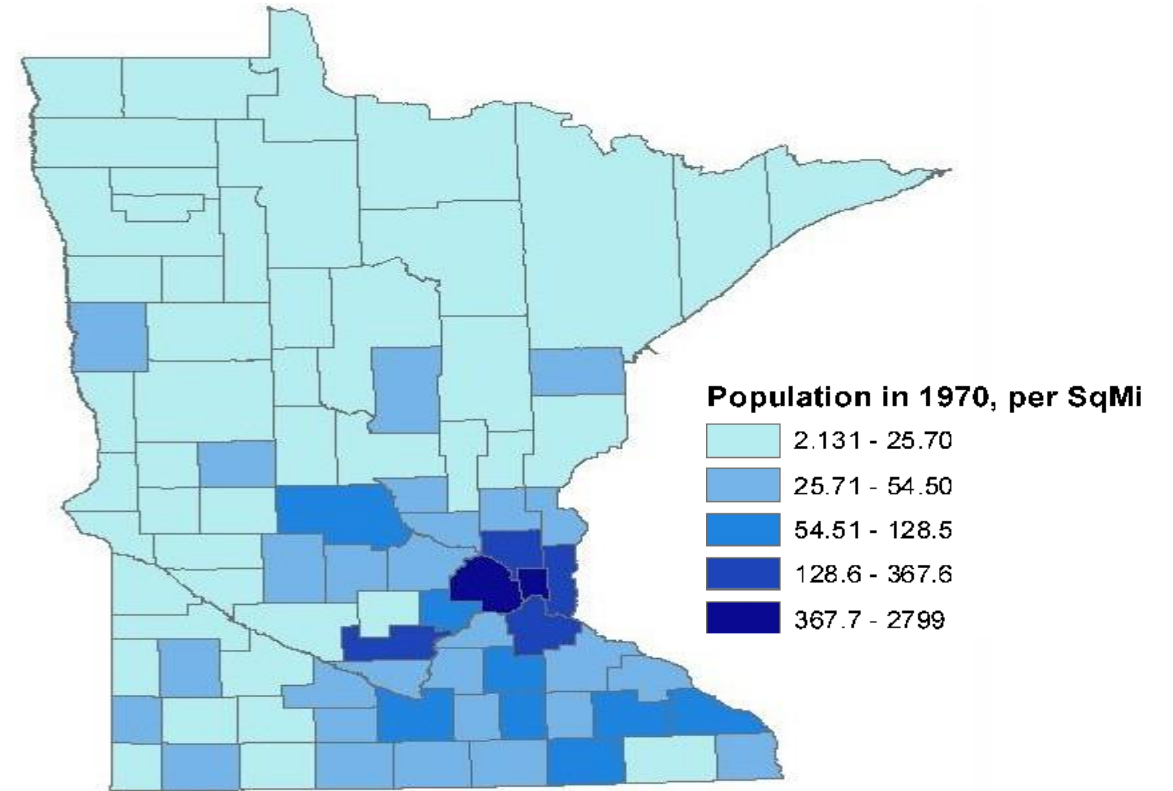
Classification

- Natural Breaks/Jenks
- Quantile
- Equal Interval Based on Range
- Equal Interval not Based on Range/Defined Interval

Natural Breaks (Jenks)

Natural Breaks has intervals that are created using natural clustering of the data. It maximizes variance between groups and minimizes variation within groups.
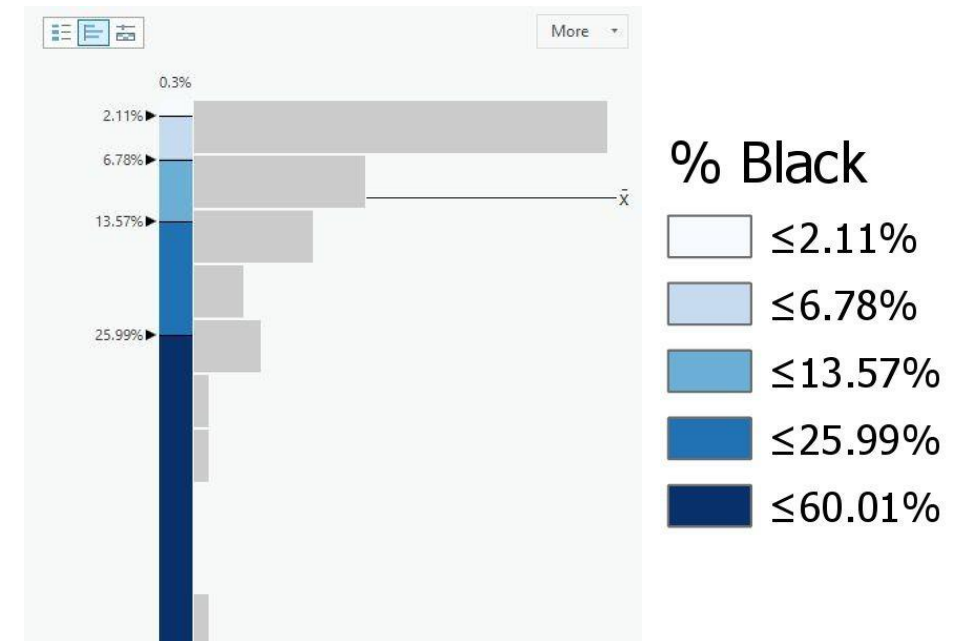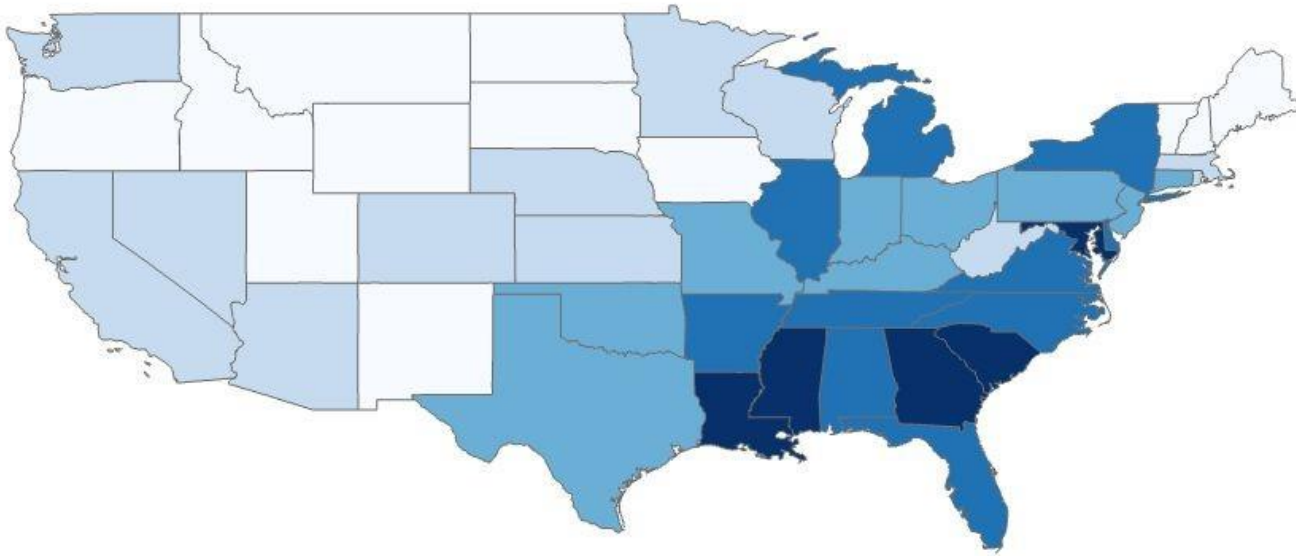
Quantile

Quantile has equal numbers of data in each class– sometimes called Quintile for 5 classes
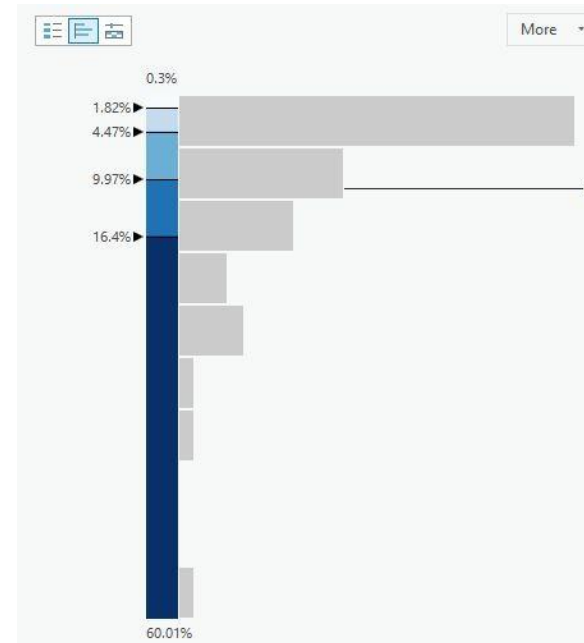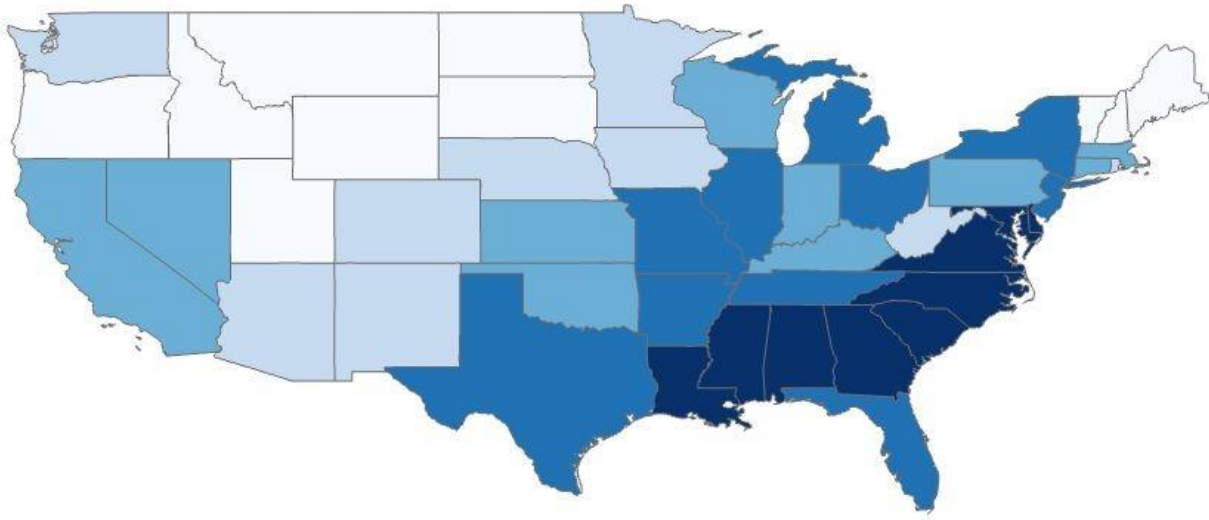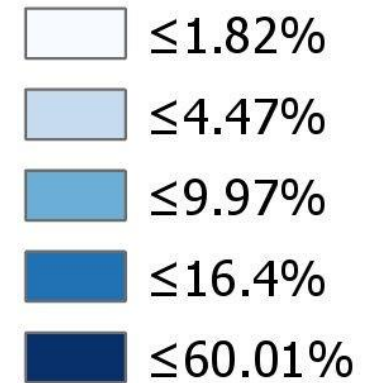
Equal Interval (Based on Range)

Equal-sized subranges

$$\frac{\text{Value of Highest Observation - Value of Lowest Observation}}{\text{Number of Classes}}$$

% Black

- ≤12.24%
- ≤24.18%
- ≤36.13%
- ≤48.07%
- ≤60.01%
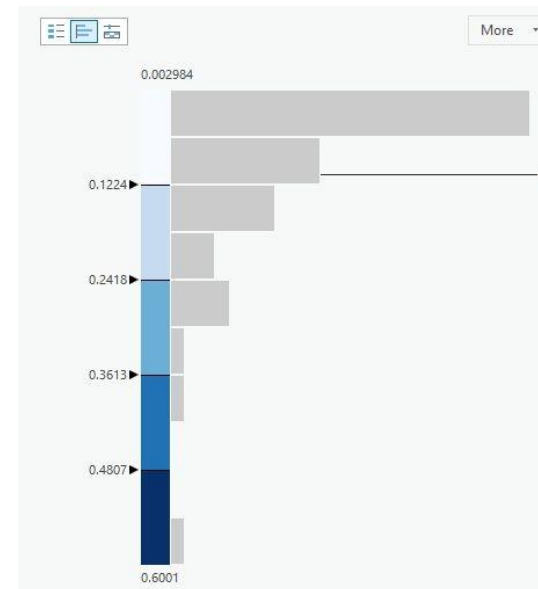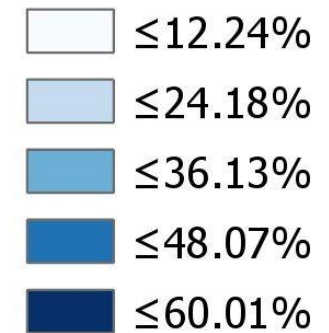
Classification: Method Comparison
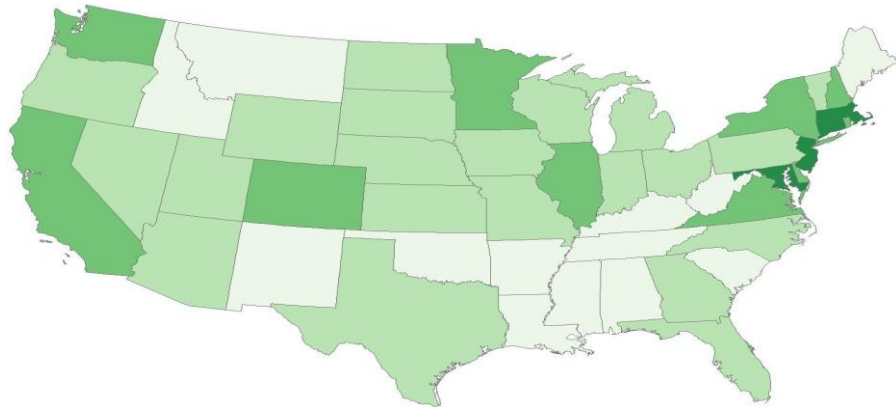
Natural Breaks
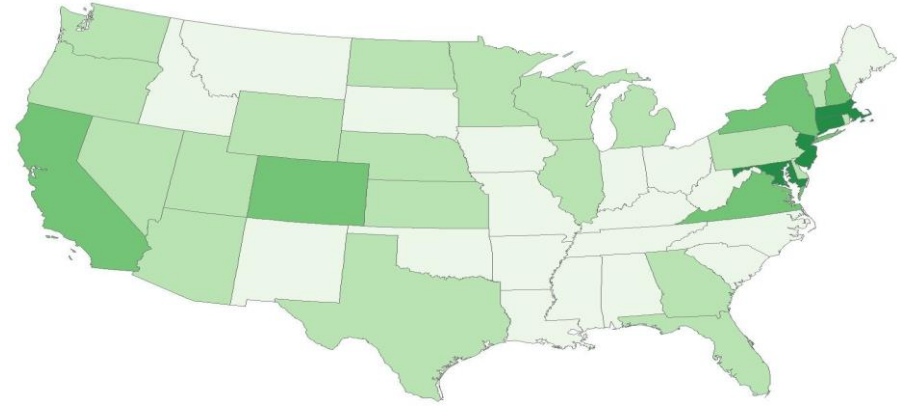
Equal Interval
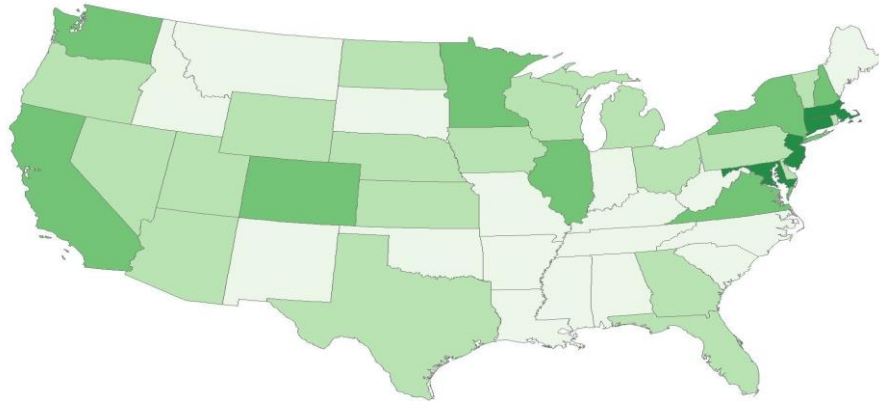
Standard Deviation
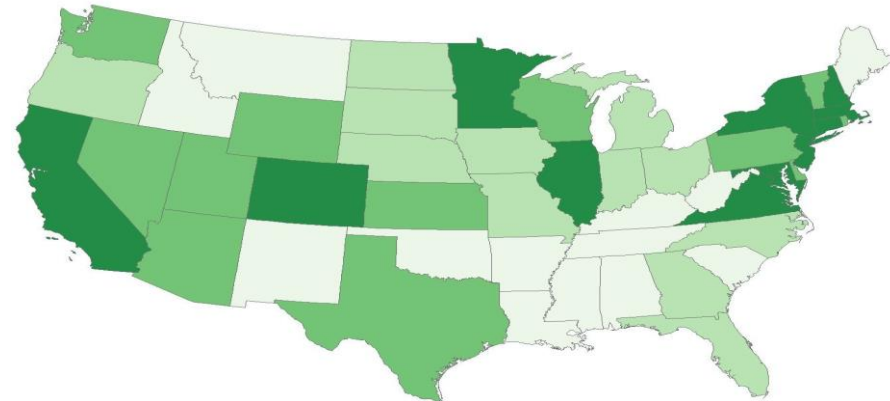
Quantile

Geocoding

Geocoding

- The process of finding the location of an address on a map.
- The location can be a pair of (X, Y) coordinate or a street address, postal delivery location, or building.
  - X – longitude (+ for E hemisphere; - for W hemisphere)
  - Y – latitude (+ for N hemisphere; - for S hemisphere)
- In GIS, geocoding requires a reference dataset that contains address attributes for the geographic features in the area of interest.

The Basics of Geocoding

X, Y Coordinates

x -3,339,584
y 3,339,584

x 1,669,792
y 1,669,792

x-axis

0,0

y-axis

x -6,679,169
y -1,669,792

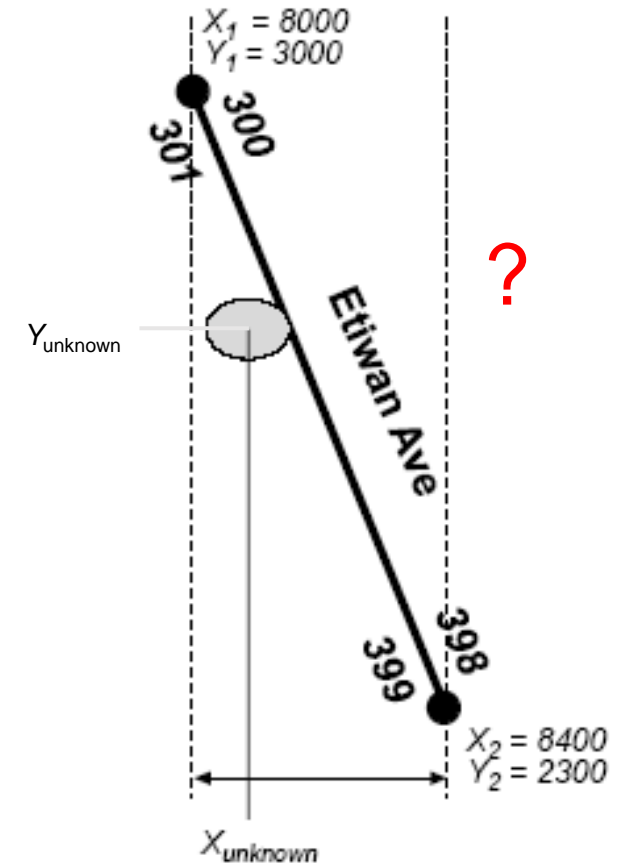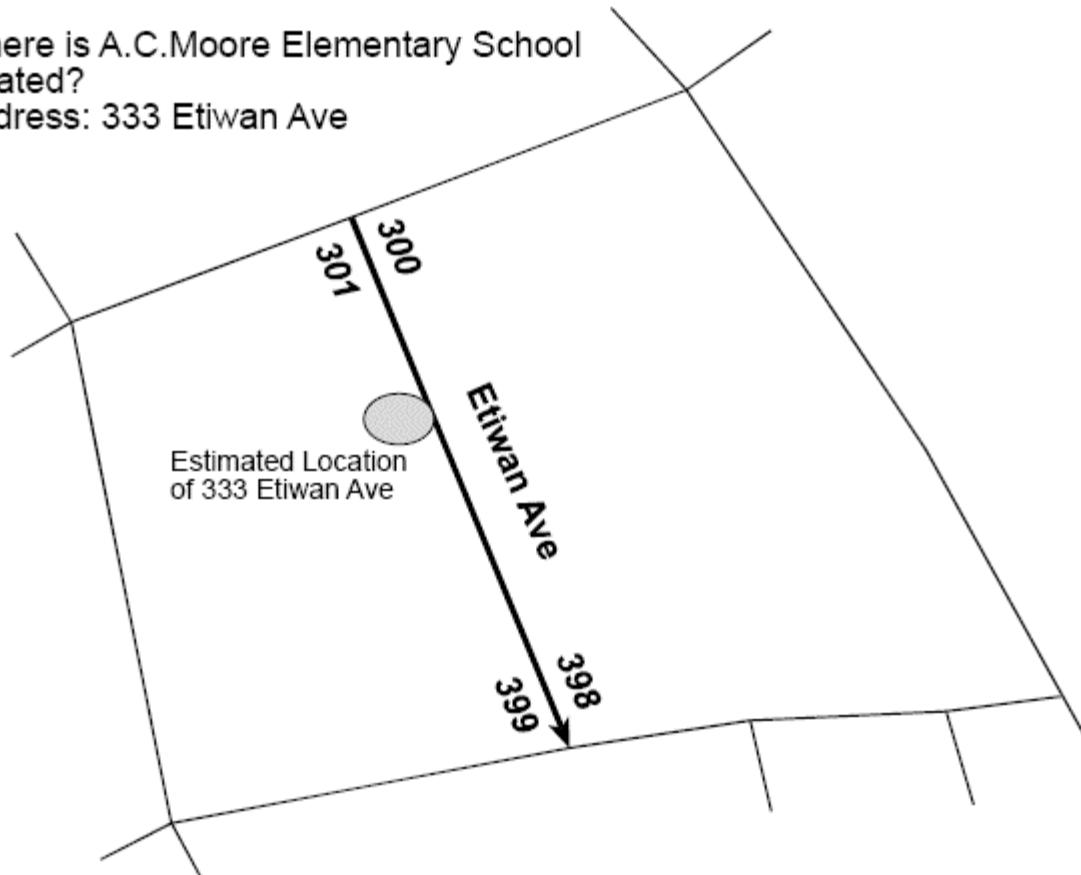x 3,339,584
y -1,669,792
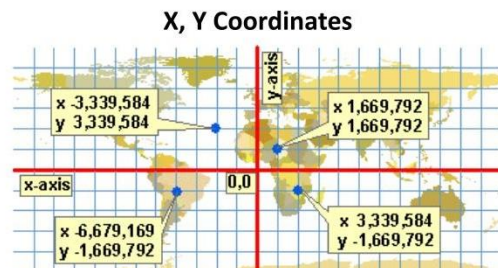
Address Matching

Address matching is the process of assigning an actual address to a geographic location on some reference files.

If an address falls within a feature's address range, it is considered a match and a location can be returned.

Address database

Reference map

Address Matching

Where is A.C.Moore Elementary School located?
Address: 333 Etiwan Ave

Estimated Location of 333 Etiwan Ave

Etiwan Ave

300
301
398
399

$X_1 = 8000$
$Y_1 = 3000$

300
301

Etiwan Ave

?

$Y_{unknown}$

398
399

$X_2 = 8400$
$Y_2 = 2300$

$X_{unknown}$

- --- Both X and Y need to be interpolated allocate the address.

RCC-GIS Geocoding Service: Formatting Data for Processing

https://gis.rcc.uchicago.edu/content/rcc-gis-geocoding-service
Based on ESRI world Geocoder

Acceptable headers:
ID
ADDRESS
NEIGHBORHOOD
CITY
SUBREGION
REGION or STATE or ST
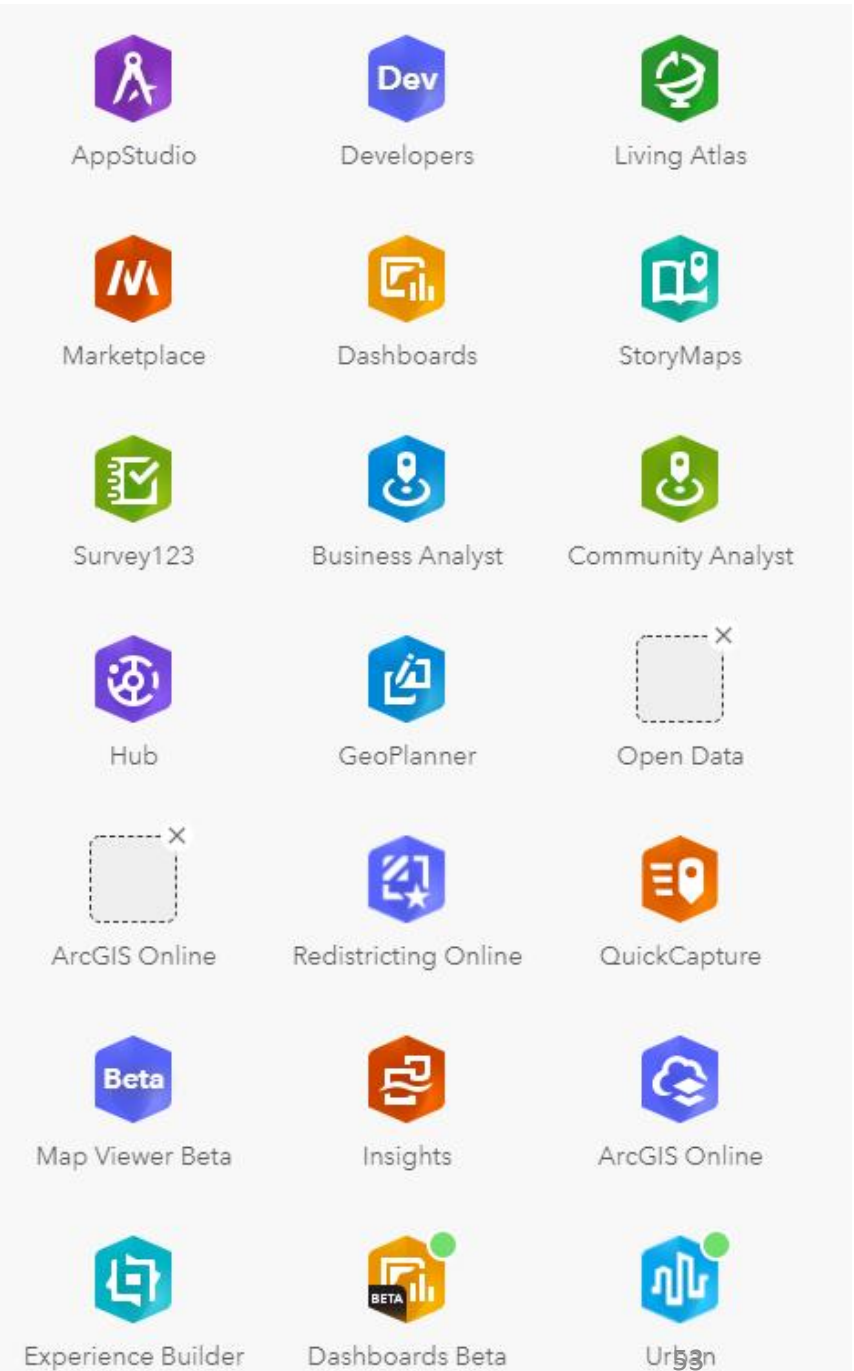POSTAL or ZIP or ZIP CODE
COUNTRYCODE

Geocoding Review

- Be careful which locator service you use, online geocoders are not HIPAA compliant

- Geocode to the appropriate geographic scale, can take care of confidentiality issues

- A high match score does not mean the point is accurate, best practice is to choose a small percentage of results to review

- Valid address does not necessarily mean correct location!

ArcGIS Online

- https://uchicago.maps.arcgis.com

- https://uchicago.maps.arcgis.com/home/organization.html#overview

- https://insights.arcgis.com/#/view/e9c2f264784543d7b9fa22c6df55cb92

Using ArcGIS Online

- Wayback App
- https://livingatlas.arcgis.com/wayback

- Navigate to https://livingatlas.arcgis.com
- Sign in using CNETid
- Search "Global Imagery Browse Services"

ArcGIS   Industries   About   Support

Parmanand Sinha
pnsinha_UChicago

ArcGIS Living Atlas of the World

Home   **Browse**   Apps   Blog   Contribute   My Contributions   My Favorites

Global Imagery Browse Services

Search examples

All   Trending ▾   Basemaps ▾   Imagery ▾   Boundaries ▾   People ▾   Infrastructure ▾   Environment ▾

Filters: All content types ⌄   All time ⌄   All regions ⌄   ☐ Esri-only content

Sort by: Relevance ⌄

21 Results

MODIS True Color - Terra Surface Reflectance

🗺 Imagery Layer By: esri

This layer provides access to NASA Global Imagery Browse Services, which delivers global, full-resolution satellite imagery. This band composition (Bands 1 4 3) most accurately represents how we see the earth's surface with our own eyes.

⊘ Authoritative

MODIS True Color - Aqua Corrected Reflectance

🗺 Imagery Layer By: esri

This layer provides access to NASA Global Imagery Browse Services, which delivers global, full-resolution satellite imagery. This band composition (Bands 1 4 3) most accurately represents how we see the earth's surface with our own eyes.

⊘ Authoritative

MODIS True Color - Terra Corrected Reflectance

🗺 Imagery Layer By: esri

This layer provides access to NASA Global Imagery Browse Services, which delivers global, full-resolution satellite imagery. This band composition (Bands 1 4 3) most accurately represents how we see the earth's surface with our own eyes.

Daily Planet Imagery

🗺 Web Map By: esri

This map shows imagery for the planet that is updated on a daily basis. It features the NASA MODIS imagery True Color band composition (Bands 1 4 3 | Red, Green, Blue) which most accurately shows how we see the earth's surface with our own eyes.
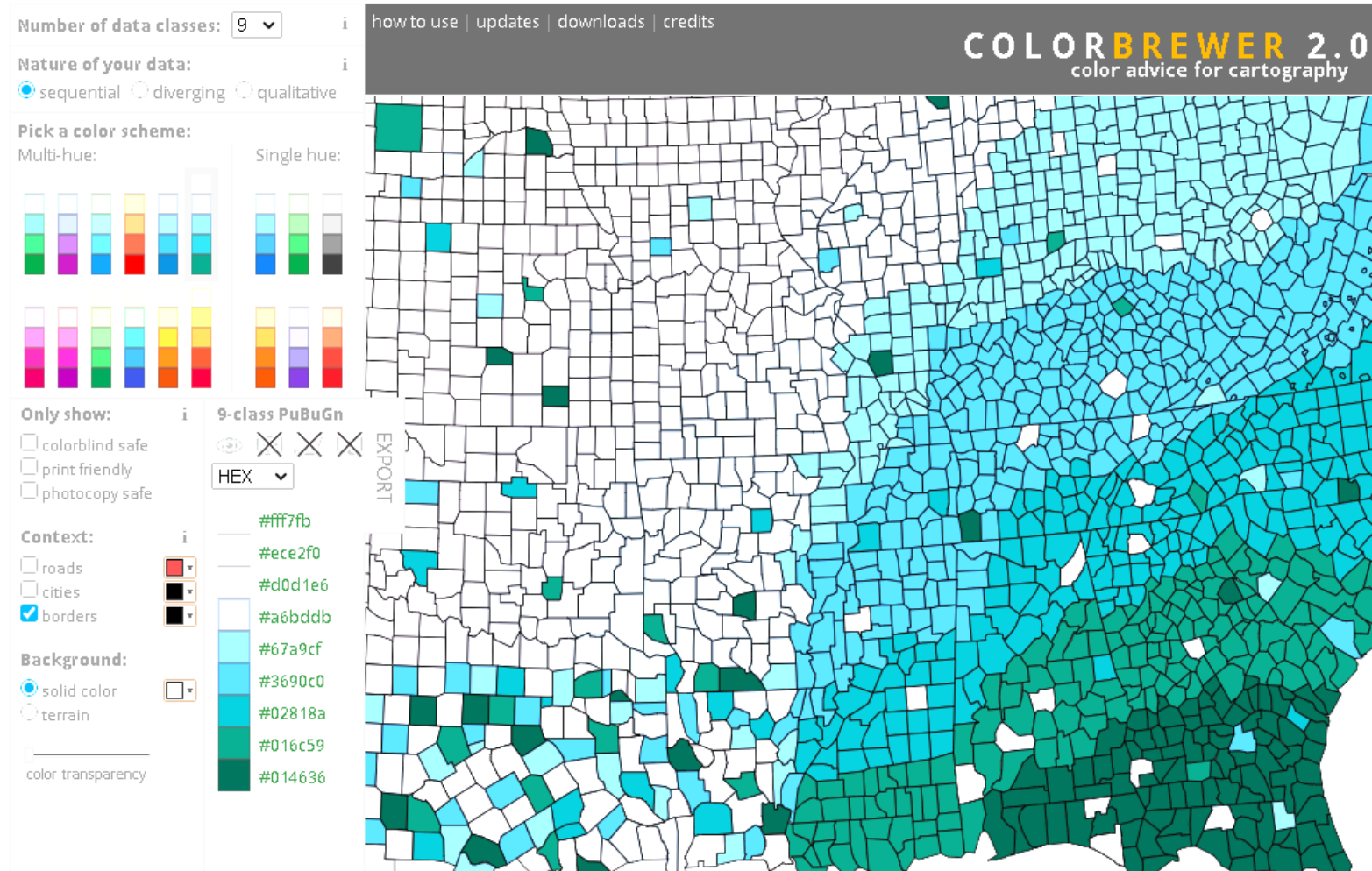
56
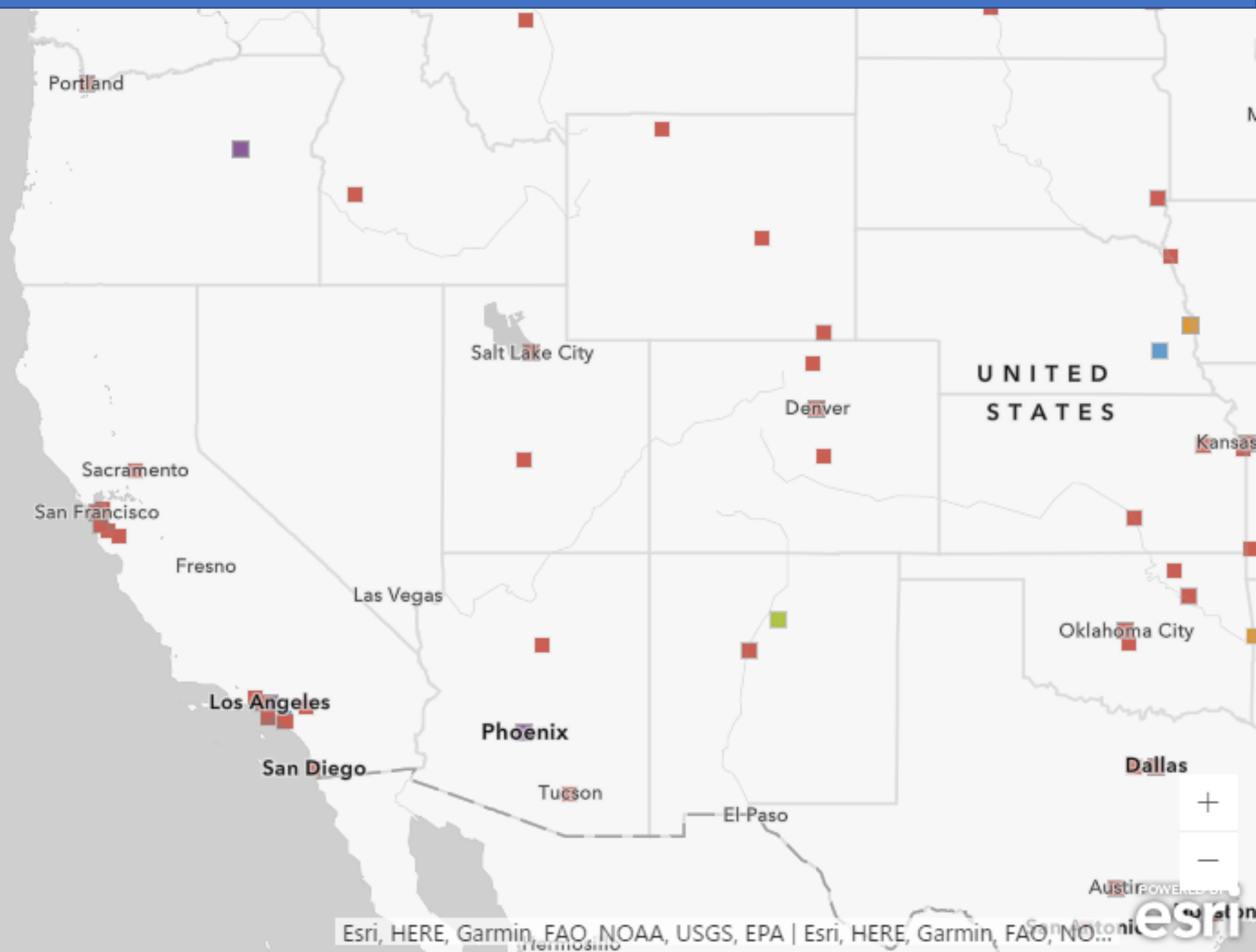
Software

QGIS

ArcGIS Online

R

ColorBrewer

MapTiler
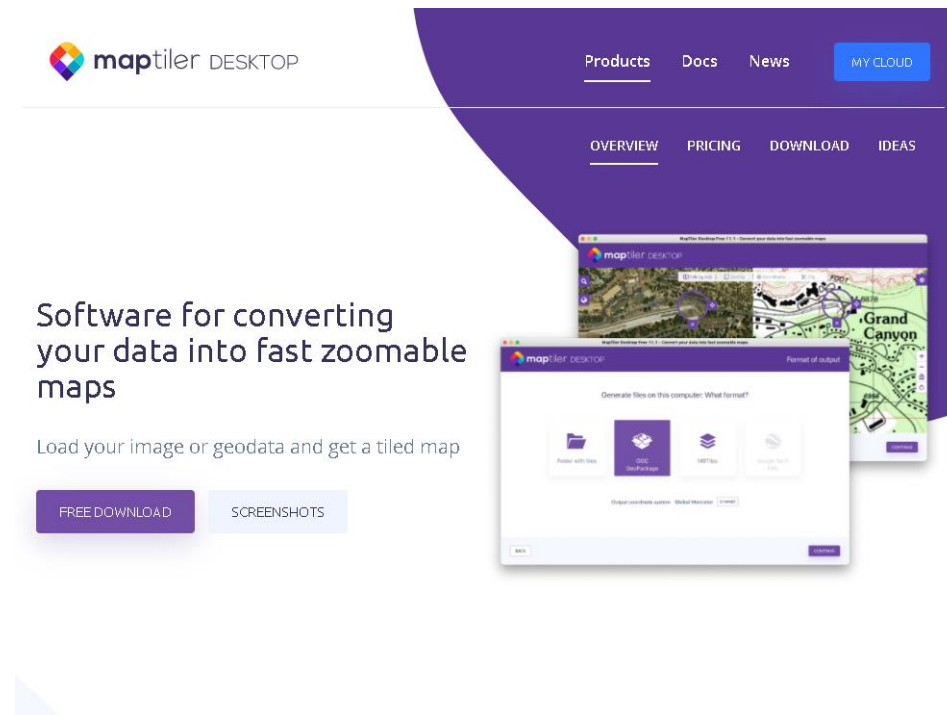
CartoDB
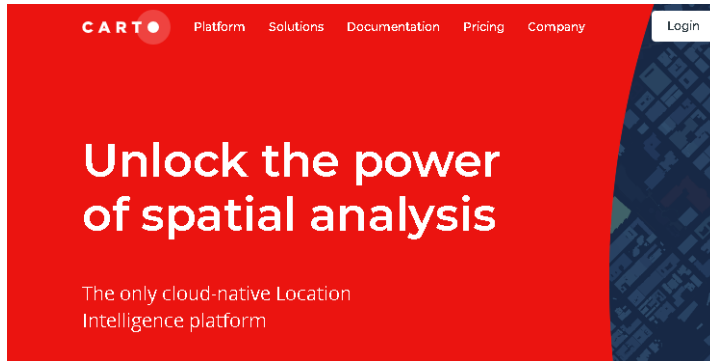
TypeID	StoreName	Address	Phone
1	Starbucks - TX - Richardson 75001	1930 N. Coit Rd, Richardson, Texas 75080	972-889-3137
1	Starbucks - TX - Dallas 75002	1380 W Campbell Rd, Dallas, Texas 75080	972-680-6010
1	Starbucks - TX - Richardson 75003	101 W. Campbell Road, Richardson, Texas 75080	972-231-1371
1	Starbucks - TX - Dallas 75004	16731 Coit Rd, Dallas, Texas 75248	214-775-0206
1	Starbucks - TX - Richardson 75005	710 Renner Rd, Richardson, Texas 75080	972-783-7665
1	Starbucks - TX - Dallas 75006	6859 Arapaho Rd, Dallas, Texas 75248	972-458-2284
1	Starbucks - TX - Dallas 75007	7995 LBJ, Dallas, Texas 75231	972-661-2666
1	Starbucks - TX - Richardson 75008	630 S Plano Rd, Richardson, Texas 75081	214-530-0183
1	Starbucks - TX - Dallas 75009	14999 Preston Rd, Dallas, Texas 75240	
1	Starbucks - TX - Dallas 75010	18208 Preston Road, Dallas, Texas 75252	972-519-0537
1	Starbucks - TX - Plano 75011	2024 15th Street, Plano, Texas 75075	972-633-5612
1	Starbucks - TX - Dallas 75012	9140 Forest Lane, Dallas, Texas 75243	214-221-0939
1	Starbucks - TX - Dallas 75013	13556 Preston Rd, Dallas, Texas 75240	972-661-8079
1	Starbucks - TX - Dallas 75014	7718 Forest Lane, Dallas, Texas 75230	214-369-3228
1	Starbucks - TX - Plano 75015	801 West 15th St, Plano, Texas 75075	972-422-5003
1	Starbucks - TX - Richardson 75016	2191 West Buckingham Road, Richardson, Texas 75081	972-238-1563
1	Starbucks - TX - Garland 75017	2645 Arapaho, Garland, Texas 75044	972-675-8853
1	Starbucks - TX - Dallas 75018	8520 Abrams Rd, Dallas, Texas 75243	214-342-6998
1	Starbucks - TX - Dallas 75019	5223 Alpha Road, Dallas, Texas 75240	972-960-9595
1	Starbucks - TX - Plano 75020	4836 W Park Blvd, Plano, Texas 75093	972-964-8190
1	Starbucks - TX - Plano 75021	1709 Preston Rd, Plano, Texas 75093	972-407-1008
1	Starbucks - TX - Dallas 75022	13350 Dallas Pkwy, Dallas, Texas 75240	972-716-0838
1	Starbucks - TX - Dallas 75023	11919 Preston Road, Dallas, Texas 75230	972-392-1680
1	Starbucks - OH 06428	650 South 3rd Street, Columbus, OH 43206	614-443-1611
1	Starbucks - OH 06429	1505 W 5th Ave, Columbus, OH 43212	614-486-0352
1	Starbucks - OH 06430	88 East Broad Street, Columbus, OH 43215	614-228-9589
1	Starbucks - OH 06431	339 North Front Street, Columbus, OH 43215	614-246-6400
1	Starbucks - OH 06432	3416 North High Street, Columbus, OH 43202	614-263-1292
1	Starbucks - OH 06433	4015 Easton Station, Columbus, OH 43219	614-337-0361
1	Starbucks - OH 06434	4784 Morse Rd, Columbus, OH 43230	614-475-4147
1	Starbucks - OH 06435	10 W Broad St, Columbus, OH 43215	614-228-4651
1	Starbucks - OH 06436	6490 Sawmill Road, Columbus, OH 43235	614-889-5914
1	Starbucks - OH 06437	1782 North High Street, Columbus, OH 43201	614-291-5687
2	Best Buy - OH001	1375 Polaris Pkwy, Columbus, OH 43240	614-430-0789
2	Best Buy - OH002	5745 Chantry Drive, Columbus, OH 43232	614-759-9829
2	Best Buy - OH	3840 Morse Road, Columbus, OH 43219	614-471-9510

Thank You

gis-help@rcc.uchicago.edy