Geocoding, Table and Spatial Join

Lecture #8 | GEOG 510
GIS & Spatial Analysis for Public Health

Varun Goel

Outline

- Geocoding
  - Addresses
  - Roads data
  - Display X,Y Data
- Table join
  - Key field
- Spatial join

Geocoding

- Address data is a form of location information
  - e.g., I could tell you my location of my apartment in x,y coordinates using my cell phone GPS… or I could give you the address
    - Addresses assume a road network
- Address information is not usable in a GIS without additional processing
  - Network datasets are used to "locate" features
  - Street name and house/building number are converted to geographic coordinates

Geocoding

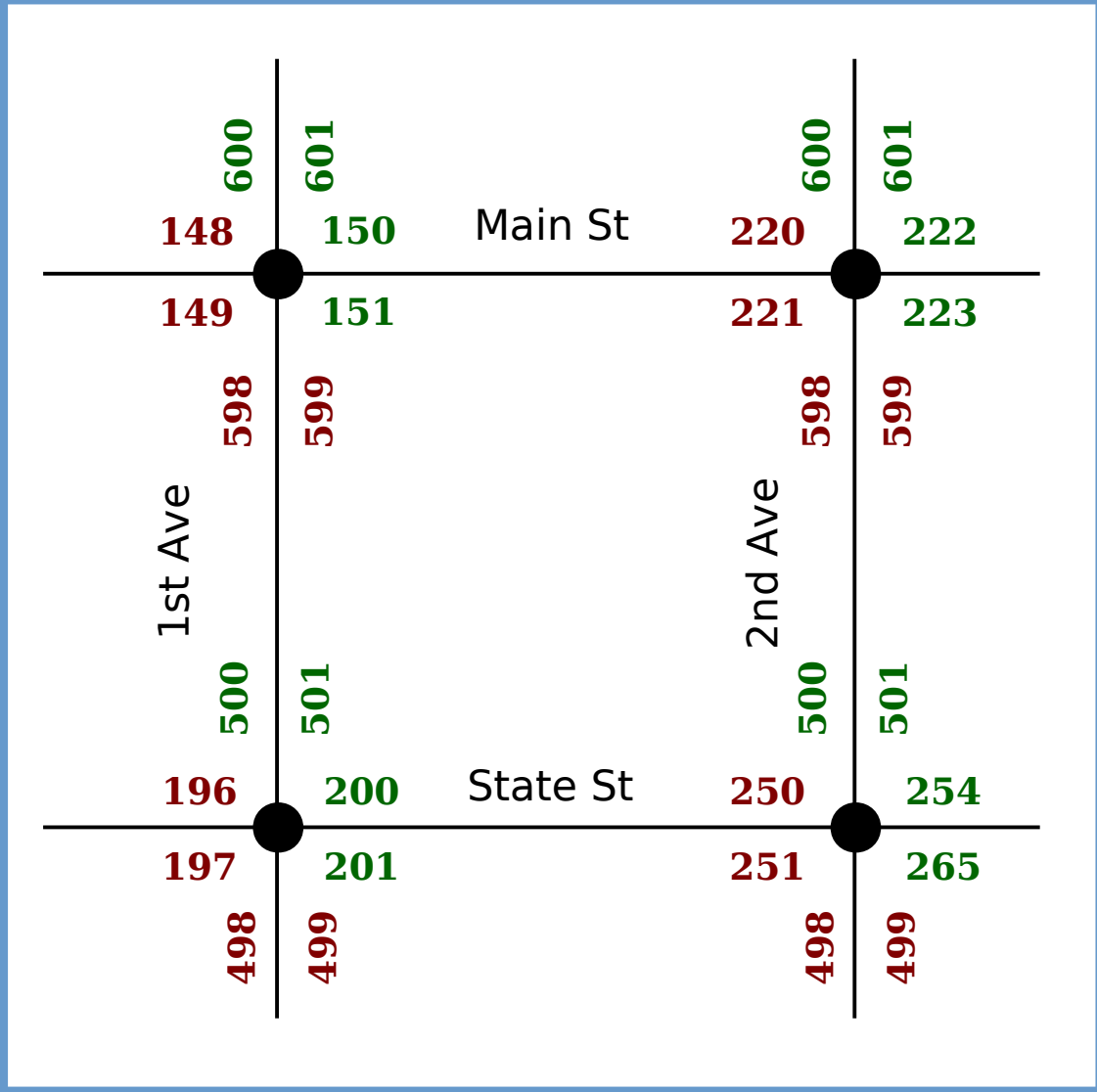
- Define location in "geographic space" using an address value
  - Also called address matching
- Requires a street/road database with specific attributes
  - Street names and address number ranges
  - Zone (in the US, a ZIP code)

Geocoding

- Attributes of the "event" are compared to the possible values in the street database
  - When/if a match is found, the event is assigned coordinate values in geographic space
    - Converts data from a list of addresses to point locations that can be mapped and analyzed
    - Very powerful tool in public health studies

Address on a Network

- A street network can store:
  - Street name
  - Address ranges associated with street segments
  - From/To addresses on left/right side of streets
    - In attribute table

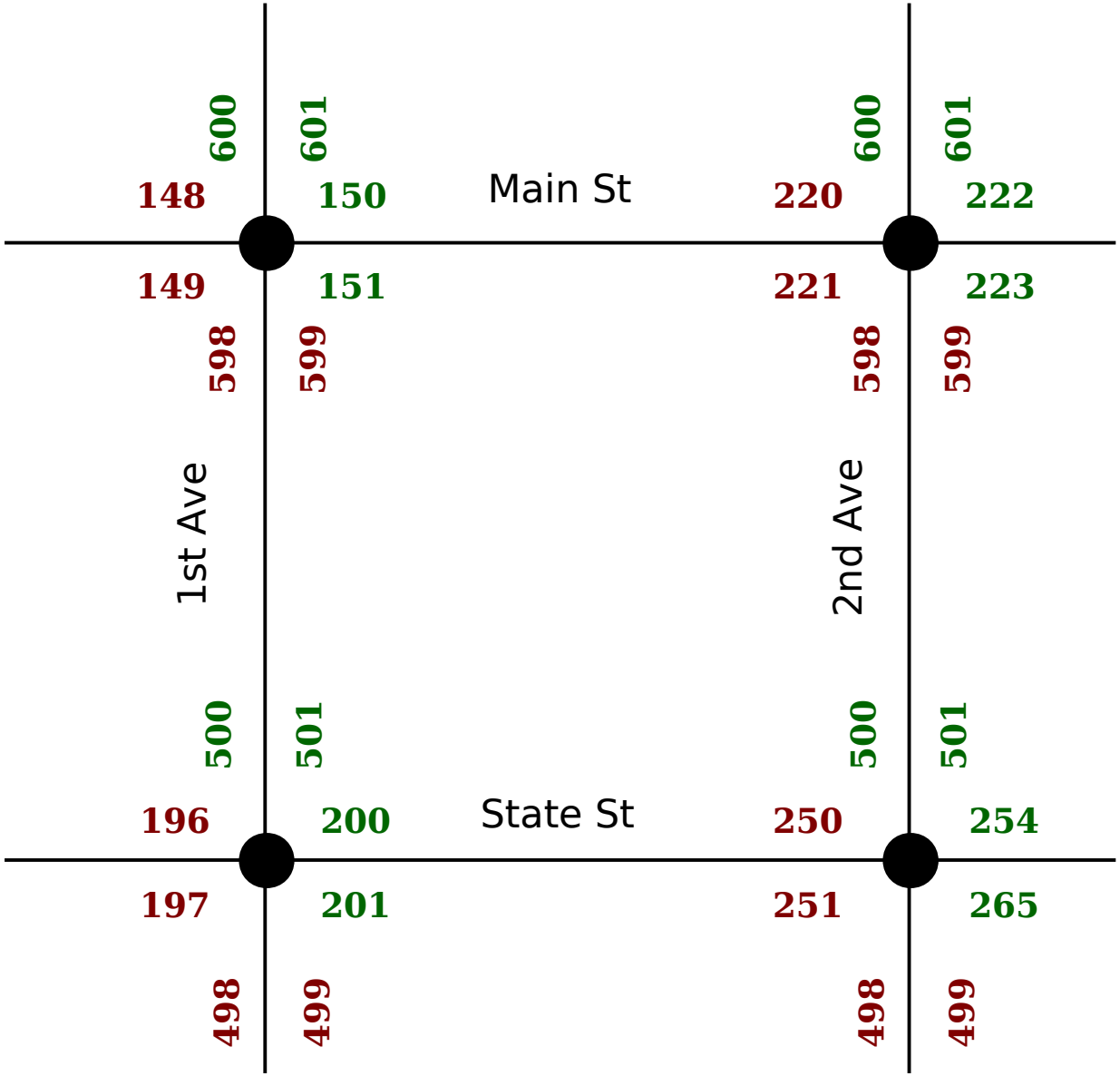


Geocoding

- An address is parsed into its elements based on specified format
  - e.g, in the address 240 State St
    - State St is the street name
    - 240 is the house number
- Once the address information has been converted to a network-like format
  - The network (table information) is searched to find features with address elements that are similar to the elements of the standardized address

Geocoding

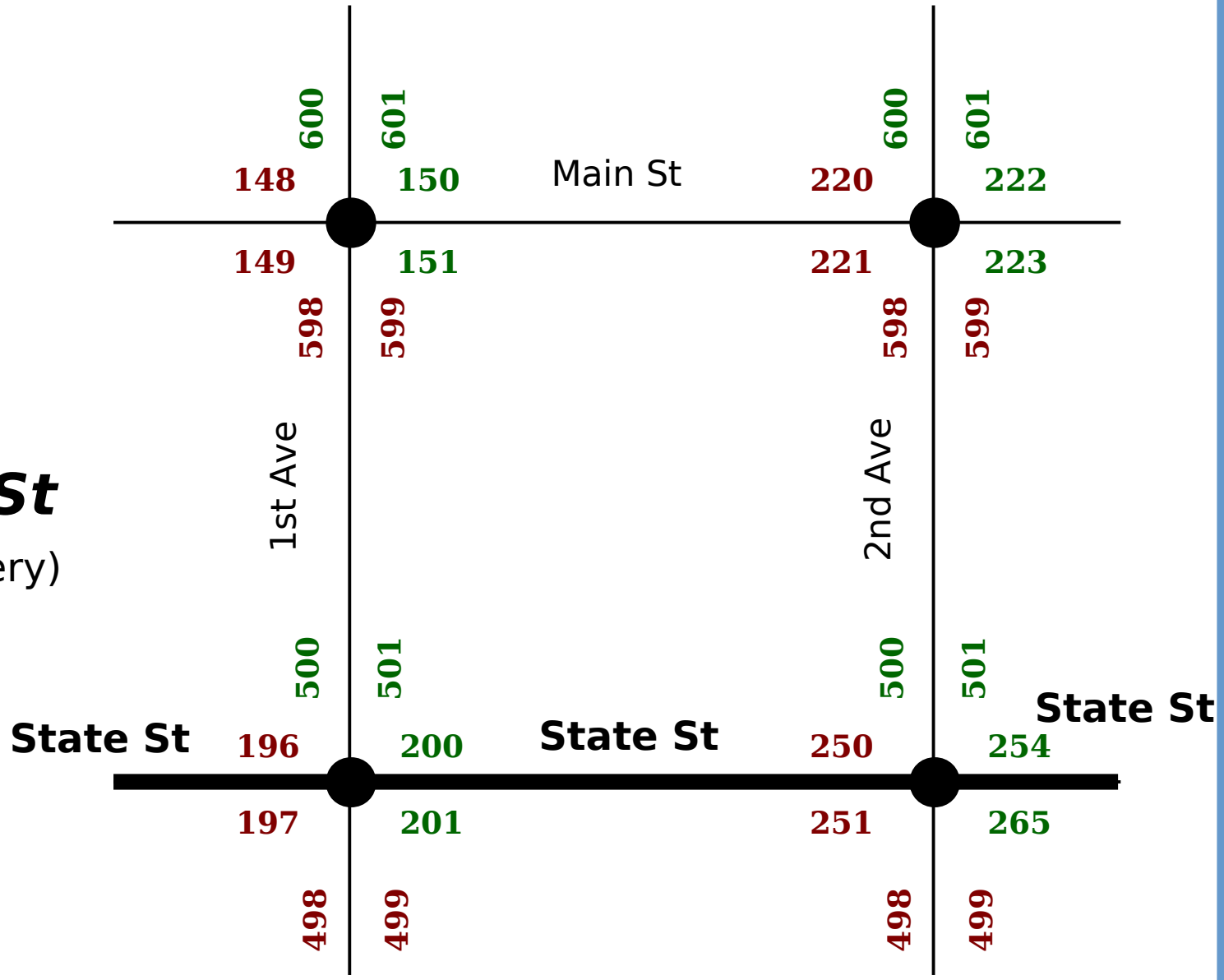
240 State St



Geocoding

240 ***State St***

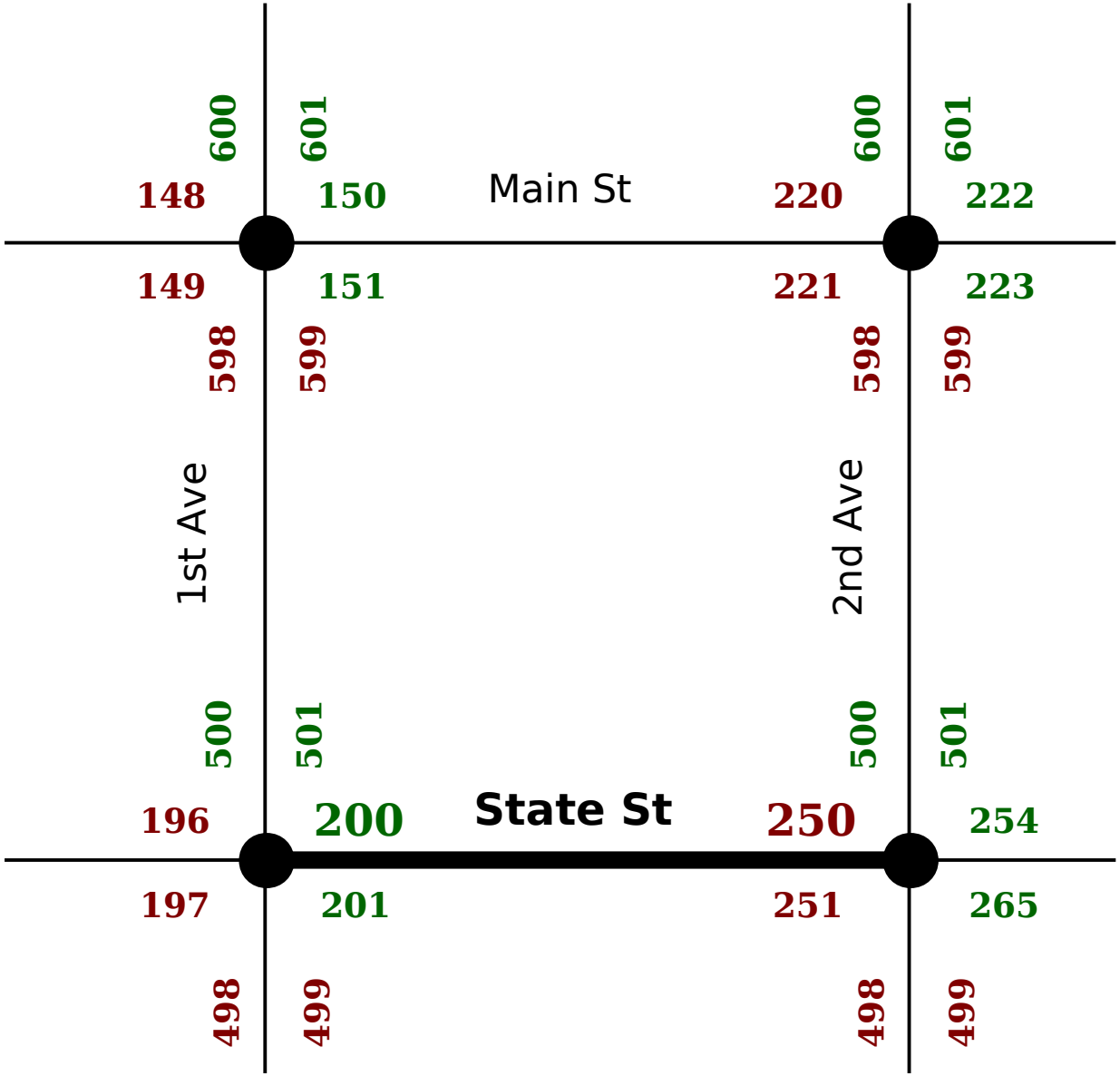
(via attribute query)



Geocoding

***240** State St*

(via attribute query)

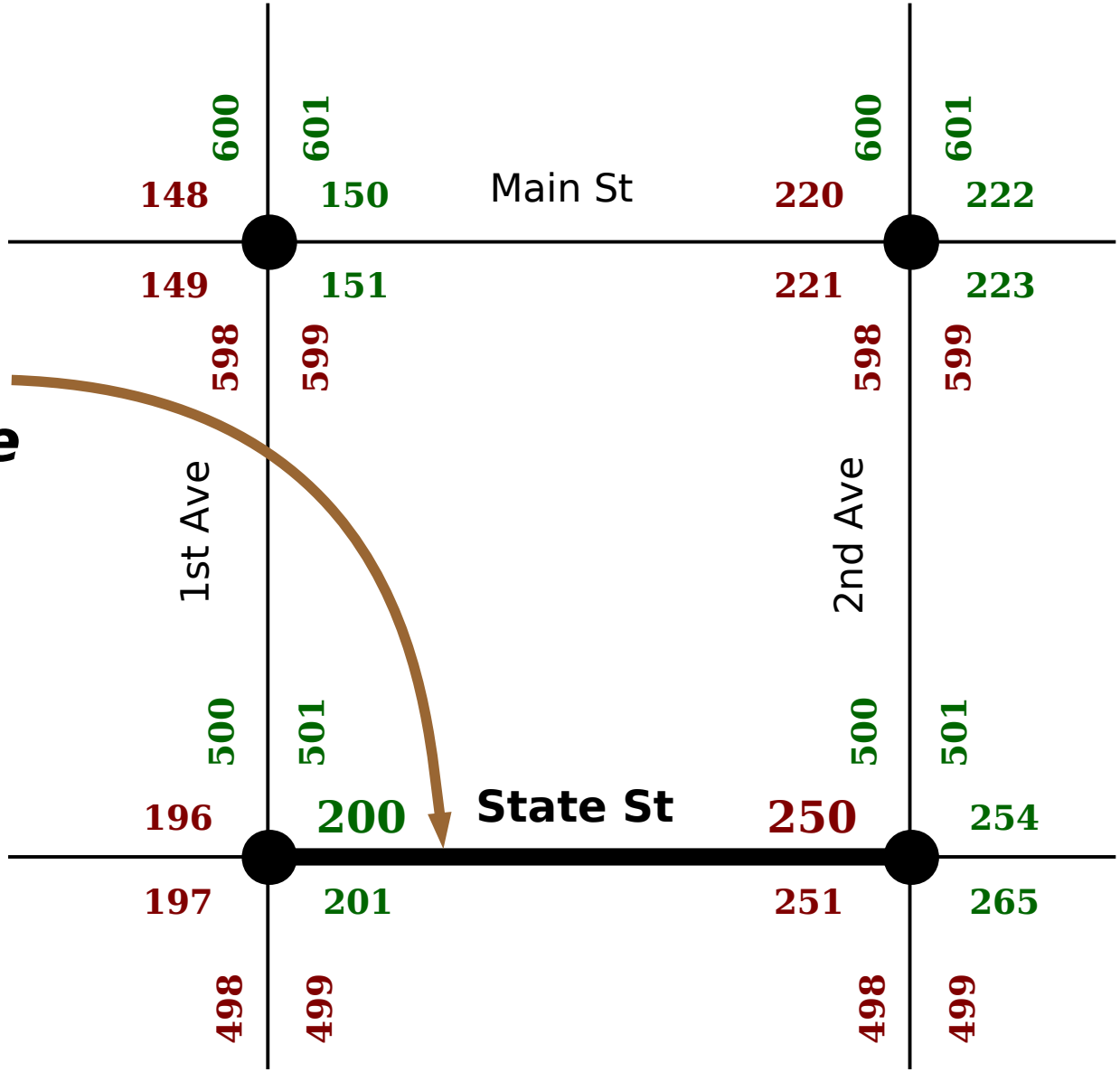


Two pieces of information
Which edge
Which side

Geocoding

240 State St

Where along this edge is the address, 240 located?

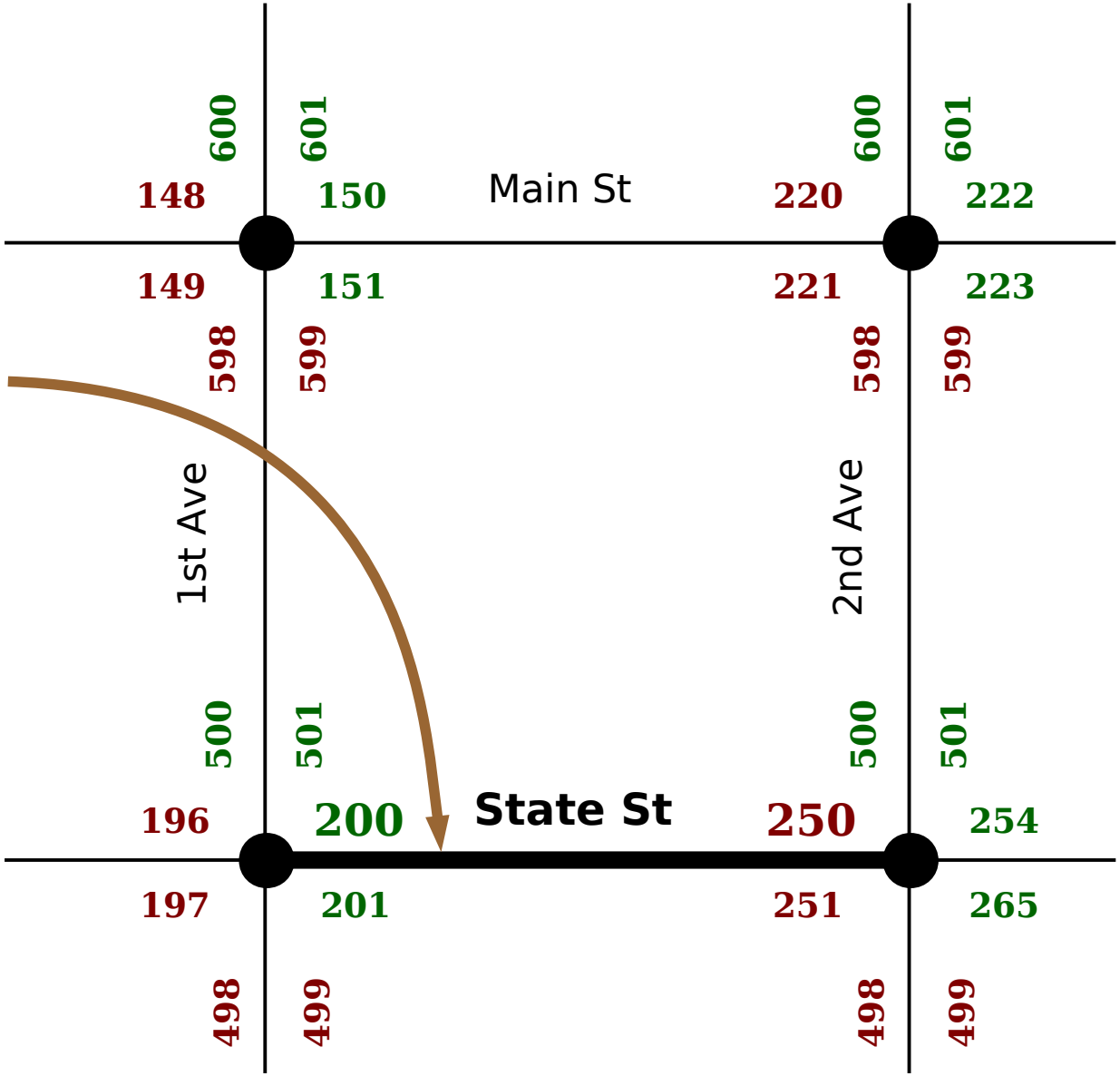


Geocoding

240 State St

This edge (line) has a “real” length, as measured from junction to junction

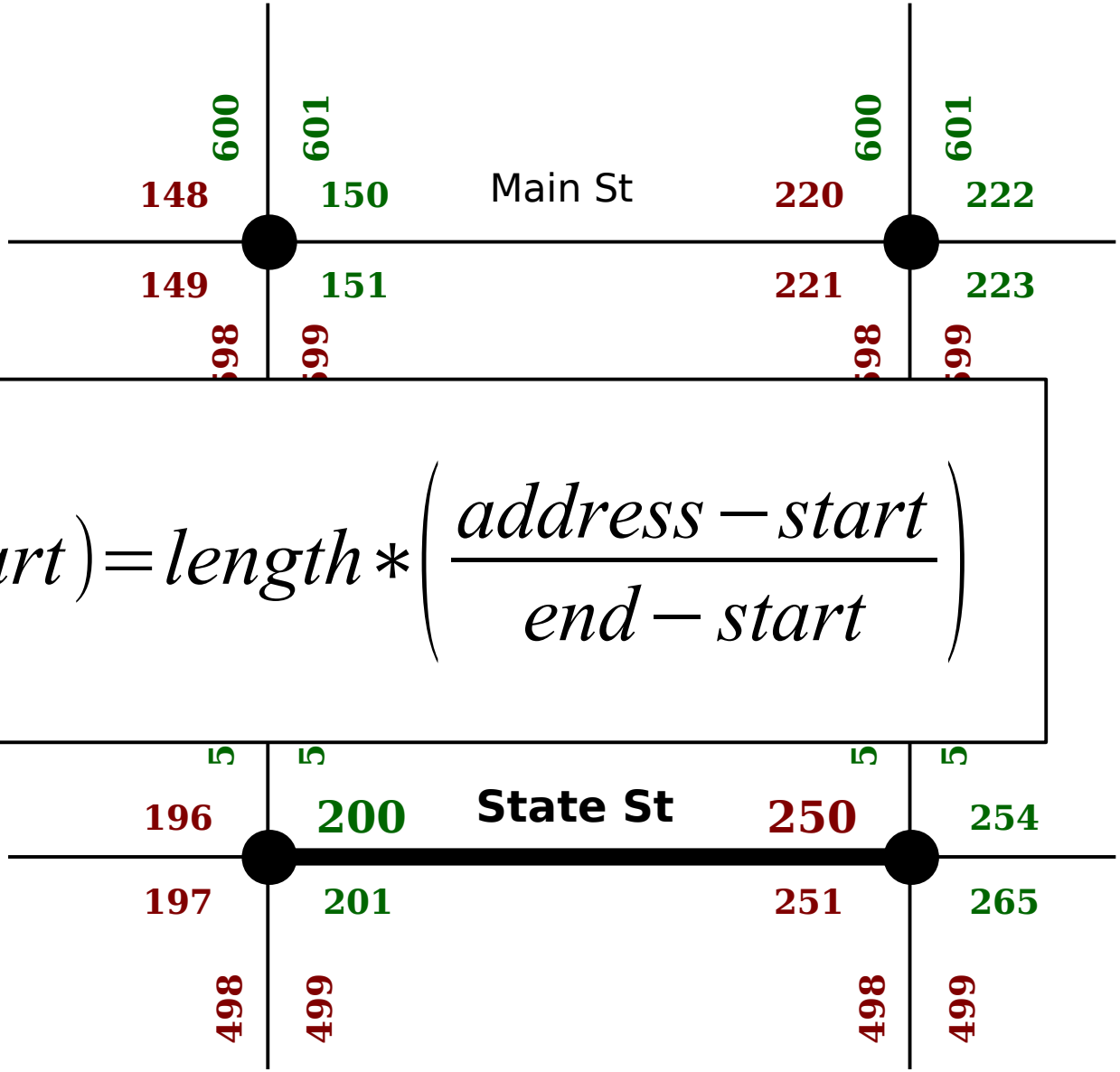
Use the edge length to locate the specific address location along the edge



Geocoding

240 State St

A simple formula!



$$distance(start) = length * \left(\frac{address - start}{end - start} \right)$$

Geocoding

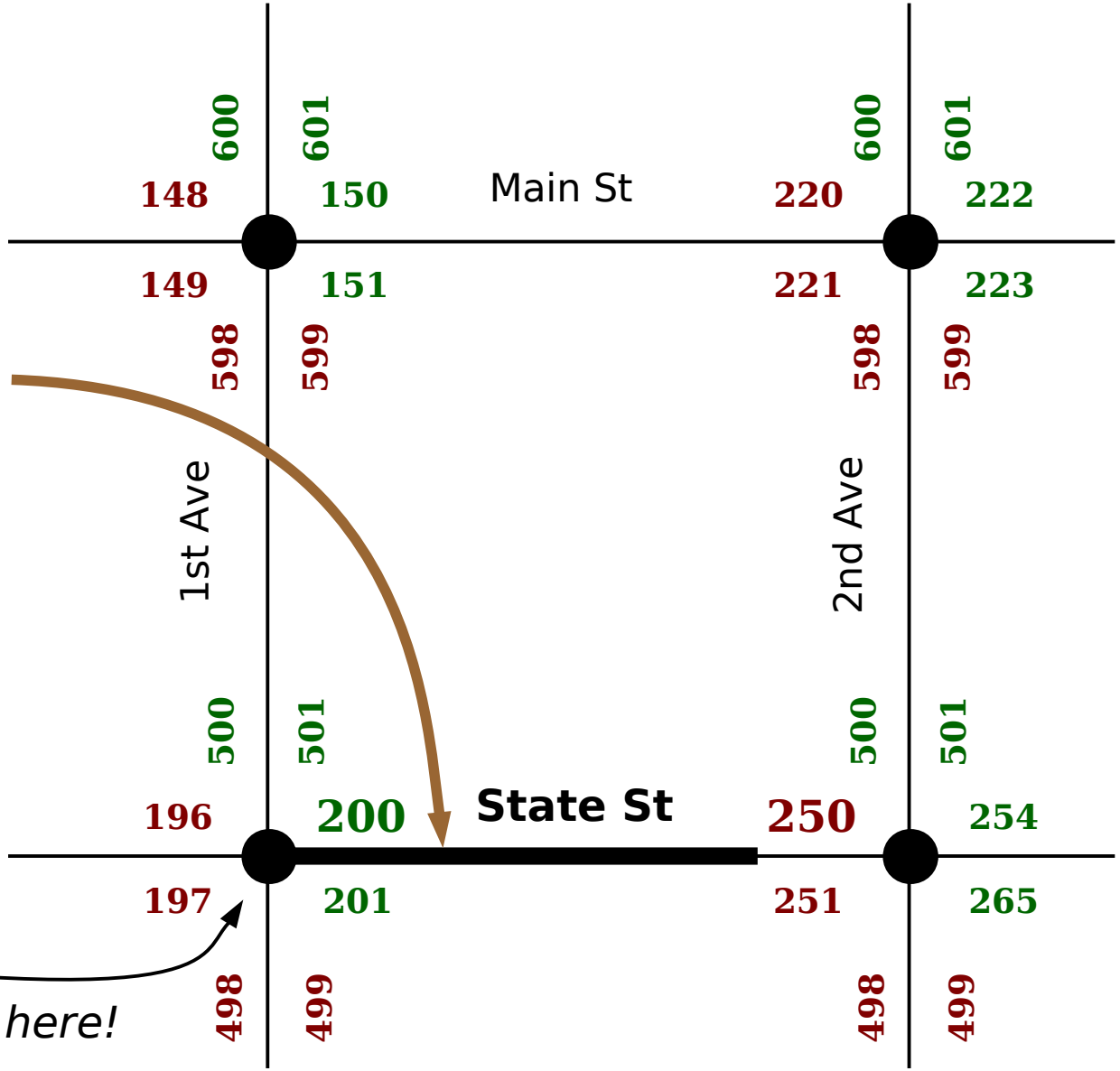
240 State St

For example: if the length of the edge is *100 meters*

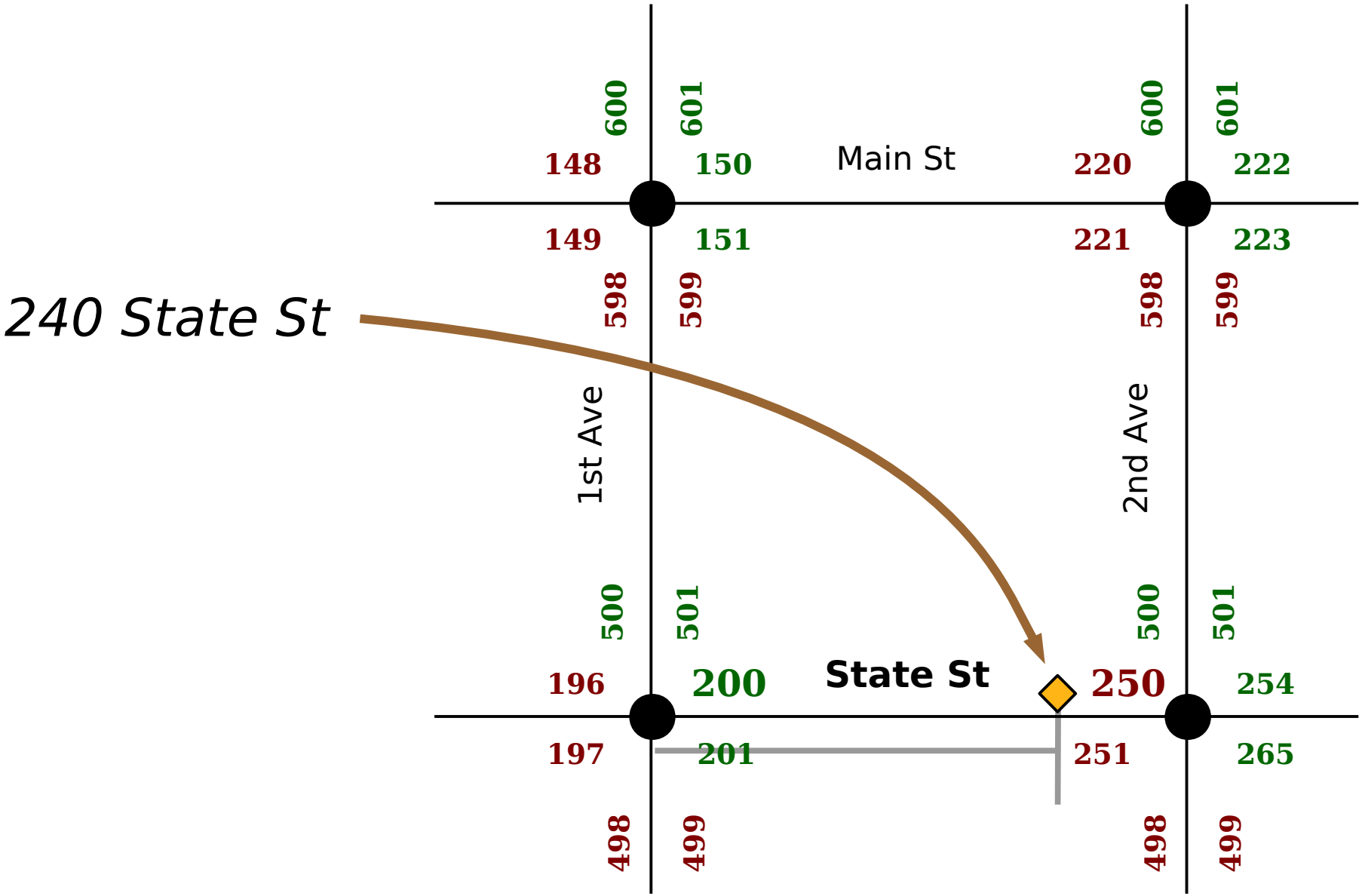
A simple formula!

$$d = 100 * \left(\frac{240 - 200}{250 - 200} \right)$$

$$d = 80$$



Geocoding


240 State St
Main St
1st Ave
2nd Ave
State St
600
601
148
150
220
222
149
151
221
223
598
599
598
599
500
501
500
501
196
200
250
254
197
201
251
265
498
499
498
499

Scoring

- A set of *candidates* that are potential matches for the address is generated
  - Based on a minimum candidate score set for the address locator
    - Common issues: 1st vs First, This St vs This Ave
- *Scores* for a candidates measure how closely each one matches the address that is being geocoded
  - Scores are adjusted based on the settings that include sensitivity to spelling mistakes

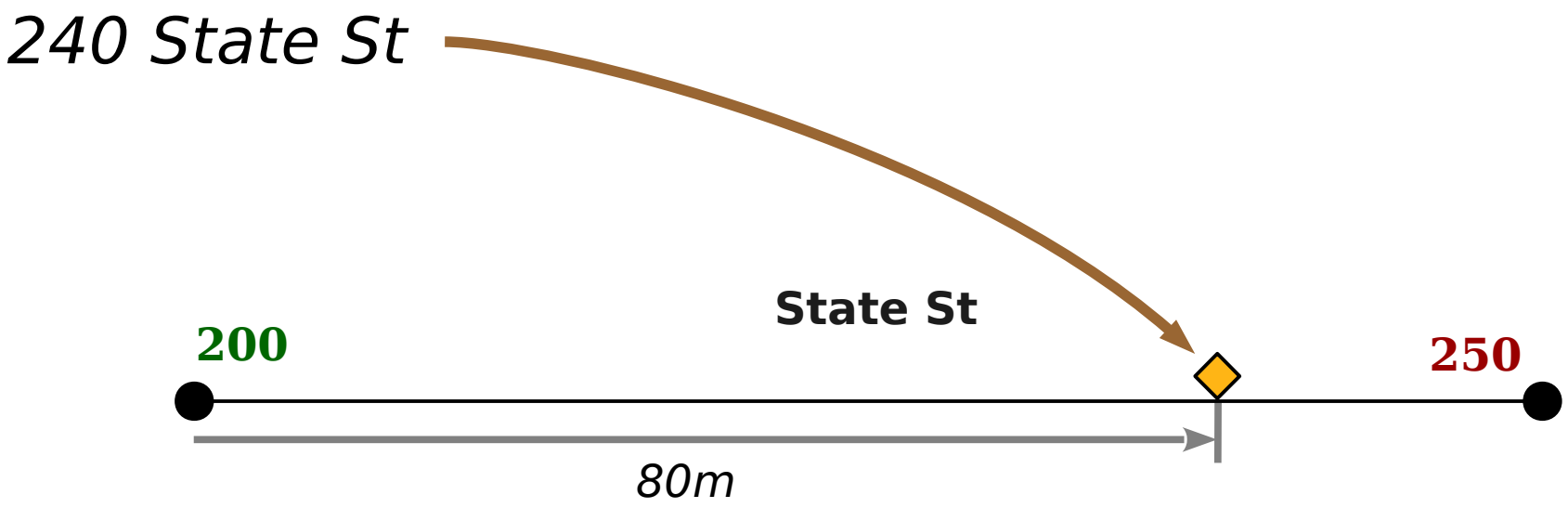
Geocoding Output

- The address locator finds the candidate with the highest score and produces the point feature
- Point features are located to the left or right of a road feature and the distance offset can be specified
  - Network Ụ Location along edge (or at junction)
  - Vector Ụ Geographic coordinates (x,y)

Geocoding

Convert to Euclidean Space from Network Space

Offset is 10 meters from edge
Offset is “left” of start location, given road attributes



Geocoding

Convert to Euclidean Space from Network Space

Offset is 10 meters from edge
Offset is “left” of start location, given road attributes

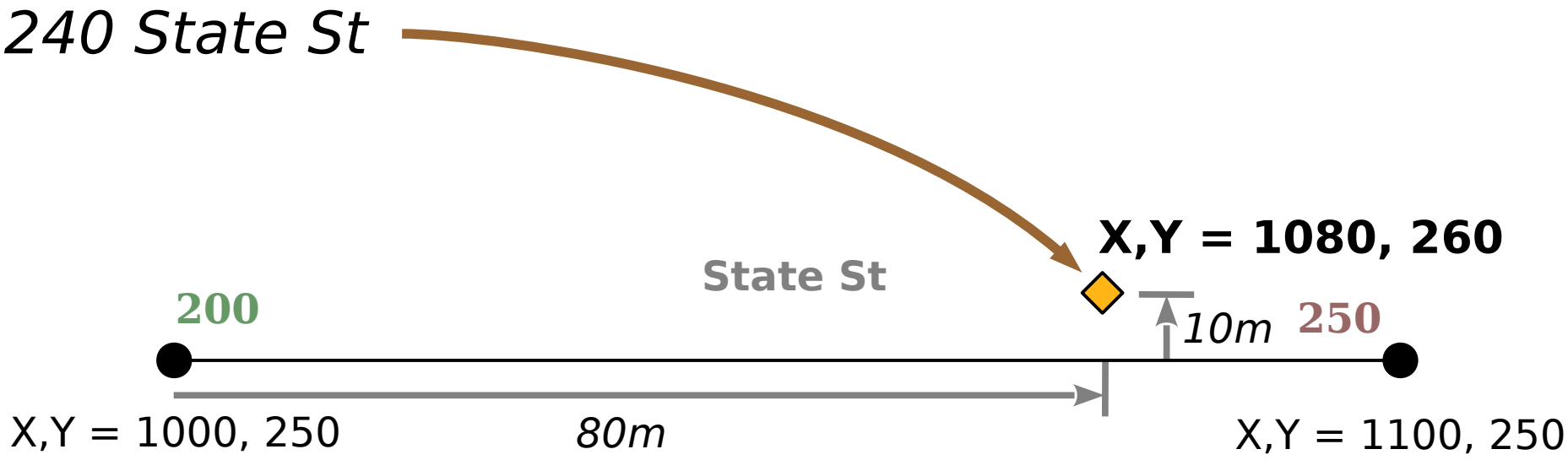
X,Y = 1000, 250

X,Y = 1100, 250

Geocoding

Convert to Euclidean Space

Offset is 10 meters from edge
Offset is “left” of start location



Geocoding Approaches

- QGIS via MMQGIS plugin
  - Accesses free online services
- ArcGIS geocoding services
  - Accesses online services
- Other online services
  - Google Maps API
  - Texas A&M

Custom R script
Format address data
Generate html call
Gather return information
Parse into x,y coordinates
Write out .csv file

Import .csv file to ArcGIS

Display X,Y Data

- Geocoding output is generally a table
  - X, Y locations of points
- Must be made into "spatial" data
  - Display X,Y Data in ArcGIS
    - Creates temporary point file

Geocoding

- Example: Bone Marrow Transplant facilities
  - Get list of Transplant facilities
    - National Marrow Donor Program
  - Copy / paste list into a text document
  - Convert to "spreadsheet-ready" format using a custom script
  - Manually edit entries into a format for geocoding

NAME	NMDP	TYPE	STREET	CITY
University of Alabama at Birmingham	NMDP Transplant Center	Bone Marrow Transplant & Cell Therapy Program	619 South 19th Street	Birmingham, AL 35249-6979
Mayo Clinic AZ & Phoenix Children's Hospital	NMDP Transplant Center	Bone Marrow Transplant Program	5777 East Mayo Boulevard	Phoenix, AZ 85054
CTI at Scottsdale Healthcare	NMDP Transplant Center		10460 North 92nd Street	Scottsdale, AZ 85258
The University of Arizona Medical Center	NMDP Transplant Center		3838 North Campbell Avenue	Tucson, AZ 85719
University of Arkansas for Medical Sciences	NMDP Transplant Center		4301 West Markham	Little Rock, AR 72205
Alta Bates Medical Center	Non-NMDP Transplant Center		2450 Ashby Avenue	Berkeley, CA 94705
City of Hope National Medical Center	NMDP Transplant Center	Hematology & Hematopoietic Cell Transplant	1500 East Duarte Road	Duarte, CA 91010-0269
Scripps Green Hospital	NMDP Transplant Center		10666 North Torrey Pines Road	La Jolla, CA 92037

Geocoding

- Example: Bone Marrow Transplant facilities
  - Use Google Geocoding API to geocode facilities
  - Import geocoded locations into ArcGIS
  - Create travel-time catchments for BMT facilities
  - Overlay population information
    - Calculate the number of people within X minutes of a BMT facility

Geocoding

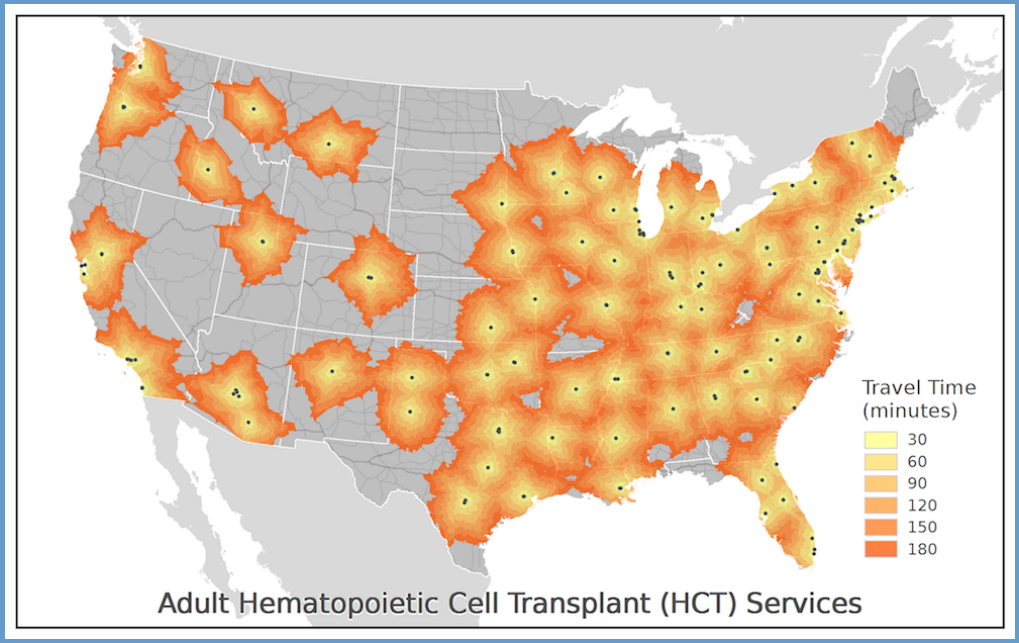
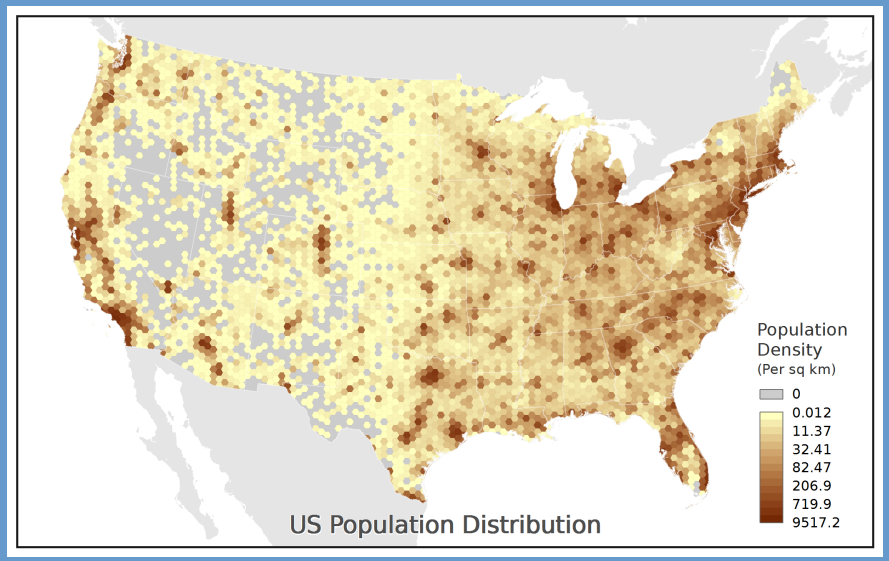


Table 2: **Summary statistics for adult and pediatric access to HCT facilities in the conterminous US.** For the access columns, the figures are in percent of the total group population. AIAN is *American Indian or Alaska Native*, HWPI is *Native Hawaiian or Other Pacific Islander*, and Multiple is *Two or more races reported.*

	Population	30 min	60 min	90 min	120 min	150 min	180 min
US Population	306,675,006	46.7	65.9	77.1	85.3	90.6	93.9
Pediatric (0–17)	73,690,271	42.4	61.0	72.0	81.0	87.2	91.5
0–9	40,274,368	42.8	61.1	72.0	81.0	87.2	91.4
10–17	33,415,903	42.0	61.0	72.1	81.1	87.3	91.5
Male	37,692,718	42.4	61.0	72.0	81.0	87.2	91.4
Female	35,997,553	42.5	61.1	72.0	81.1	87.2	91.5
White	39,578,416	32.9	55.3	68.1	78.7	86.0	90.7
Black	10,352,121	55.4	68.3	78.1	86.7	93.7	96.9
Hispanic	17,070,811	52.3	66.8	75.1	81.7	85.4	89.6
AIAN	613,648	18.5	28.6	36.9	48.3	59.1	68.0
Asian	3,087,832	69.7	82.8	89.0	92.9	95.1	96.8
HWPI	95,339	50.5	67.6	78.3	83.6	86.2	88.3
Other	222,302	53.7	72.2	80.9	88.3	93.1	95.6
Multiple	2,669,802	43.3	62.7	74.1	82.9	88.7	92.3
All minorities	34,111,855	53.5	67.7	76.6	83.8	88.6	92.3
Adult (18+)	232,984,735	48.0	67.4	78.6	86.6	91.7	94.7
18–29	51,416,023	50.1	68.3	79.3	86.9	91.9	94.8
30–44	60,627,621	50.8	70.2	80.5	87.7	92.3	95.1
45–59	64,223,419	47.3	67.5	78.8	86.7	91.7	94.7
60–74	38,277,320	43.6	63.6	75.9	85.0	90.6	94.0
75+	18,440,352	44.4	63.7	76.1	85.4	91.0	94.2
Male	113,037,737	47.5	67.0	78.3	86.4	91.5	94.5
Female	119,946,998	48.5	67.8	78.9	86.9	91.8	94.8
White	156,474,473	41.1	63.2	76.3	85.6	91.3	94.5
Black	27,291,874	62.7	75.7	85.0	91.8	96.2	98.0
Hispanic	33,246,692	59.3	74.6	80.9	85.5	89.1	92.4
AIAN	1,528,071	27.4	41.4	52.8	63.4	72.8	78.7
Asian	10,826,539	75.7	88.7	92.6	95.0	96.4	97.7
HWPI	250,796	55.0	74.5	82.0	86.8	89.9	92.6
Other	378,964	67.5	82.8	88.3	91.9	94.8	96.5
Multiple	2,987,326	53.4	71.7	81.4	88.0	91.9	94.5
All minorities	76,510,262	62.0	76.2	83.5	88.8	92.5	95.0



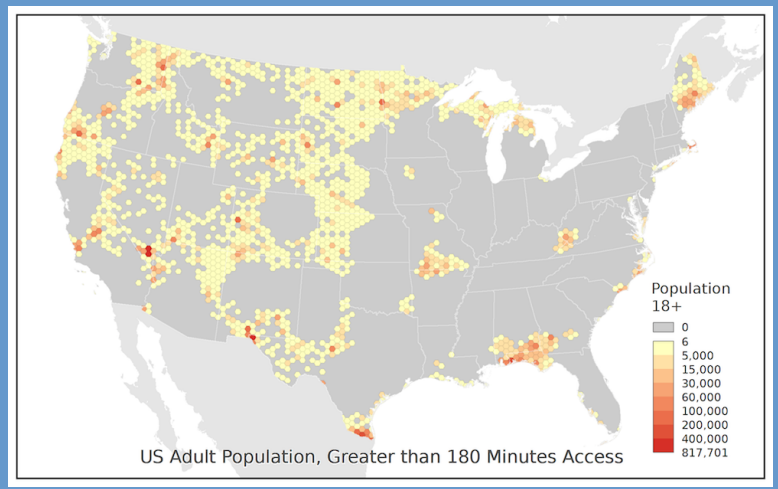


Table Join

- Often, available health data is in a non-spatial data format
  - Tables!
    - Tabular data may be joined or linked to spatial features (foundation data)
    - Requires a common field (key field)
    - Requires a "unique identifier" for each feature or object
      - e.g., Name, ID, code
      - Not always easy… unfortunately

Table Join

- Key fields
  - **Numeric IDs** are the best
  - Best to stay away from using **names**
    - Spelling
      - "St. XXX" won't join to "Saint XXX"
    - Whitespaces
      - "Chapel Hill" won't join to "Chapel Hill "
    - Capitalization
      - "CHAPEL HILL" won't join to "Chapel Hill"

Table Join

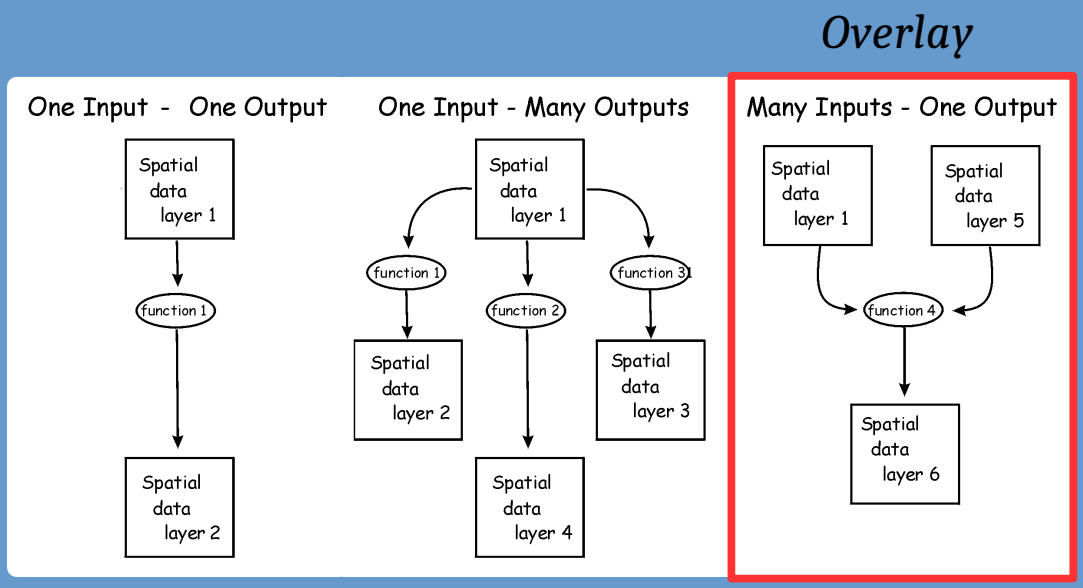
- A key field is not always available...
  - ...in the spatial data
  - ...in the tabular data
- You may have to construct/create a field in one or both tables
  - See Table Join ref document
- Key fields must be of the same data type
  - e.g., numbers stored as text can cause issues

PA Schools Example

- PA Vaccination/Exemptions
  - Data
    - Schools file with vaccination Information
    - Public Schools file with addresses
    - Private Schools file with addresses
  - Process
    - Get list of schools with vacc info
    - Subset Pub / Priv Schools file
    - Geocode
    - Table join

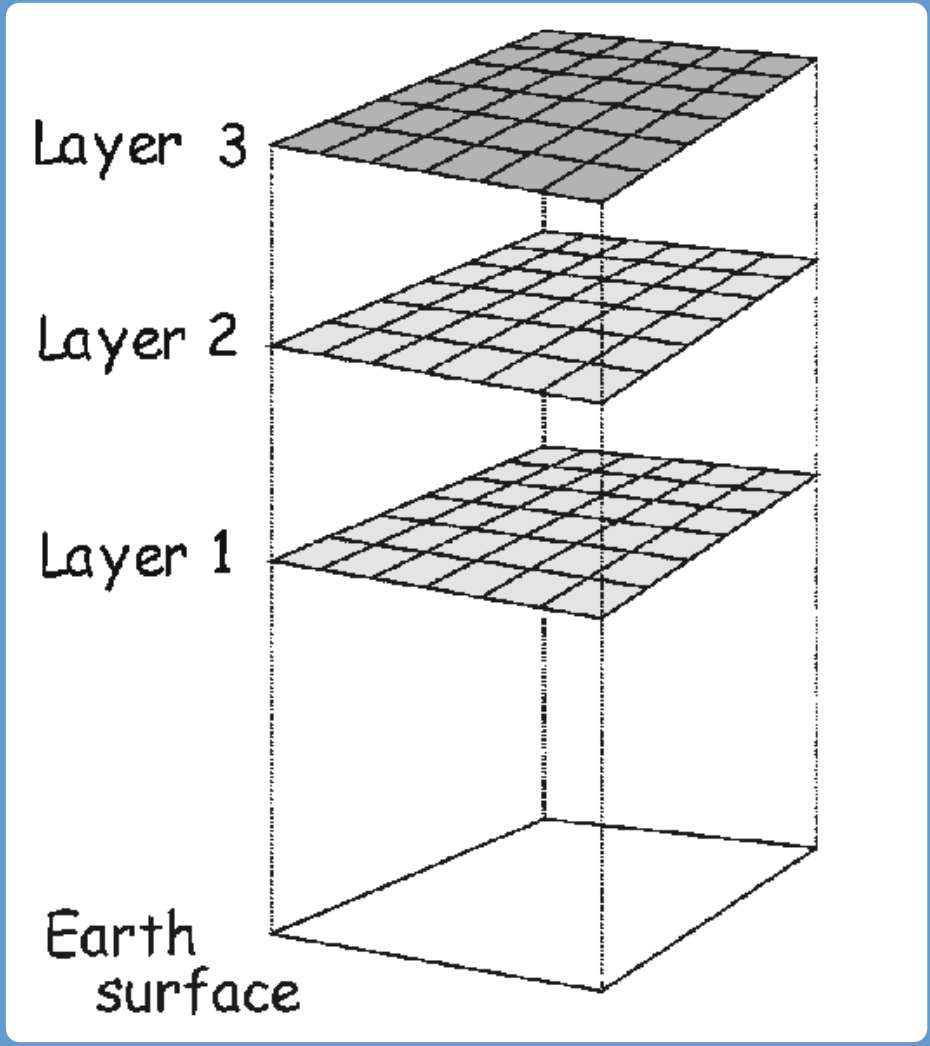
Overlay

- Overlay functions combine information from two or more layers
- Combinations are all based on some shared location
  - Functions are *explicitly spatial*
  - However, the usefulness of the output is generally related to both location and attributes



Data Layers in GIS

- Information is organized in data layers
- Each layer represents a single "theme"
- Overlay functions leverage the "location" information in each layer



Spatial Join

- Spatial Join is a *vector* overlay operation that transfers attributes from one data layer to another based on spatial location
  - Features must already be "spatial"
    - Instead of a common *attribute* (field)
    - Table join
  - Creates a new data layer
    - Spatial features are the same, but have "new" attributes appended to them

Spatial Join

- Spatial Join is a *vector* overlay operation that transfers attributes from one data layer to another based on spatial location
  - How attributes are transferred is dependent on…
    - The two types of input features (points, lines, polygons)
    - The selected approach of combining attributes from multiple features

Spatial Join

- What about raster data?
  - Assign values in a raster layer to point locations
  - Lines and polygons are a bit more complex...
    - Conversion might be necessary

Keywords

- Geocoding
  - Scoring
  - X,Y Data
- Table join
  - Key field
- Spatial join