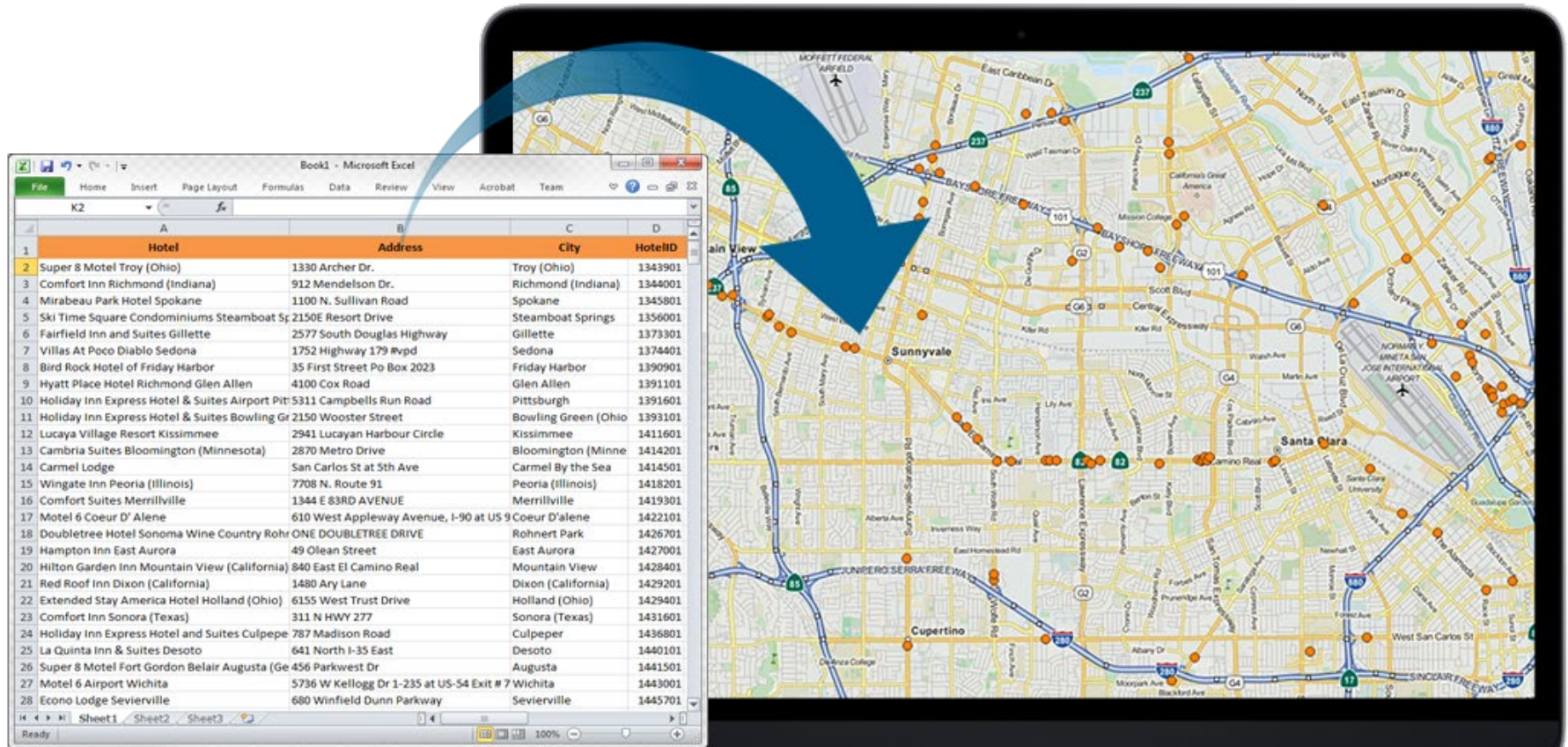# Introduction to Geocoding

May 21, 2019

What is Geocoding?

Geocoding uses a description of a location, most typically a **postal address** or place name, to find **geographic coordinates** from spatial reference data such as building polygons, land parcels, street addresses, postal codes, etc.

Topics

1. TIGER Basics

2. Geocoding vs. Address Matching

3. Online Geocoders

4. Geocoding Errors
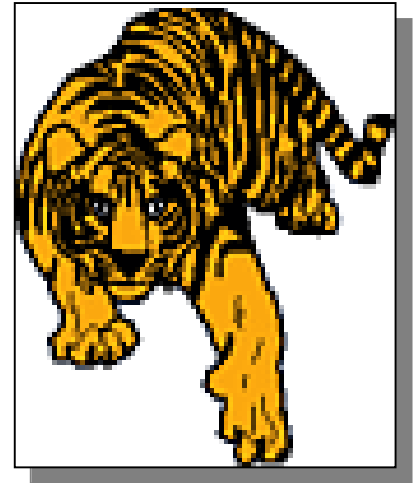
Hands on Training

1. RCC Geocoder

2. Google Geocoder API

3. OSM Geocoder

4. Comparison

TIGER = Topologically Integrated Geographic Encoding and Referencing

Initial TIGER database created in 1990

TIGER Basics

- A system for recording the location of geographic features associated with U.S. Census.

- The system includes the location and attributes of census related point, line and area features.

What information is stored?

- Census Feature Class Codes (CFCC)
  - Road = A, Transport terminal = D
- Census geography (Federal Information Processing Standard- FIPS)
  - E.g. State, county, census tract
- Feature name
- Address range
- Source material

	Attributes of tgr45079lkA												_	□	×
	TLID	FEDIRP	FENAME	FETYPE	FEDIRS	CFCC	FRADDL	TOADDL	FRADDR	TOADDR	ZIPL	ZIPR	SOURCE	CENSL
	71807274		Zion Church	Rd		A41	1042	1044	1043	1045	29016	29016	B	0
	71807276		Zion Church	Rd		A41	1041	1001	1040	1000	29016	29016	B	0
	71807672		Zion Church	Rd		A41	1046	1098	1047	1099	29016	29016	B	0
	71828518		Zion	Ave		A41	1600	1610	1601	1611	29201	29201	A	0
	71828519		Zion	Ave		A41	1612	1658	1613	1699	29201	29201	A	0
	71828521		Zion	Ave		A41	1660	1710	1701	1721	29201	29201	A	0

Record: |◄ ◄ | 0 | ► ►| Show: | All | Selected | Records (0 out of 26924 Selected.) | Options ▼

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
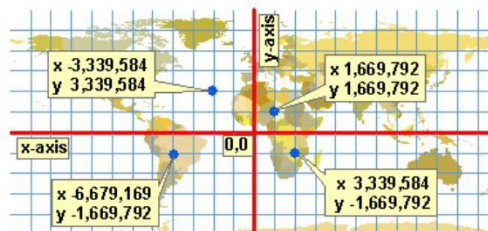
Address Matching

The process of relating two or more data files using a common address field.

A process that compares an address or a table of addresses to the address attributes of a reference dataset to determine whether a particular address falls within an address range associated with a feature in the reference dataset.

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

Address database

125 E. Somewhere

1000 N. Else Street

19234 W. Hwy 6

9938 Cat Place

455 Main Road SW

Address Matching

Reverse Address Matching

Main Road SW

W Hwy 6

Reference map

Address Matching

What is address matching?

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

Common reference files

- A **point/polygon** reference file where each feature contains *1 discrete address location*

  E.g. Parcel or parcel centroids

- A **point/polygon** reference file where each feature contains *a series of address locations*

  E.g. Zip + 4

- A **line** reference file where each feature contains *address ranges*

  E.g. TIGER roads (streets)

Table 8.1 Commonly used road segment attributes and how they are named in TIGER 2000, TIGER 2014 Address Range Feature County, and The National Map transportation layers.

Attribute	TIGER 2000	TIGER 2014	The National Map
Prefix direction of road (N. Smith St.)	FEDIRP	n/a	n/a
Name of road (N. **Smith** St.)	FENAME	FULLNAME	Full_Street_Name
Type of road (N. Smith **St.**)	FETYPE	n/a	n/a
Suffix direction of road (Canal Street **E.**)	FEDIRS	n/a	n/a
Start of address ranges on left side of road	FRADDL	LFROMHN	Low_Address_Left
End of address ranges on left side of road	TOADDL	LTOHN	High_Address_Left
Start of address ranges on right side of road	FRADDR	RFROMHN	Low_Address_Right
End of address ranges on left side of road	TOADDR	RTOHN	High_Address_Right
Zip code on left side of road	ZIPL	ZIPL	Zip_Left
Zip code on right side of road	ZIPR	ZIPR	Zip_Right
Code to determine the kind of road (residential, highway, etc.)	CFCC	ROAD_MTFCC	Road_Class

Parsing Addresses – breaking up street addresses into component pieces, i.e., prefix, number, name, type, and suffix

Address Matching

- Attribute query, Topological structure of street data (i.e. direction), Attributed with to and from address range

- Linear interpolation – a method used in geocoding to place an address location among a range of addresses along a street segment.

Attribute query of address characteristics
--- city, State zip code
--- street name, type

Direction based on street number
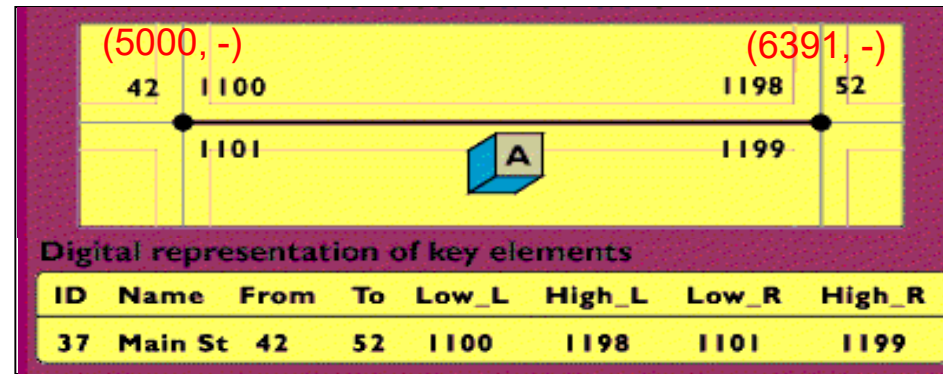
Left → 1100 to 1198

Right → 1101 to 1199

1147 is the odd number → **RIGHT**

Attributed with to and from address range

1147 is between 1101 to 1199

Locating 1147 Main Street on map

ID	Name	From	To	Low_L	High_L	Low_R	High_R
37	Main St	42	52	1100	1198	1101	1199

Linear Interpolation

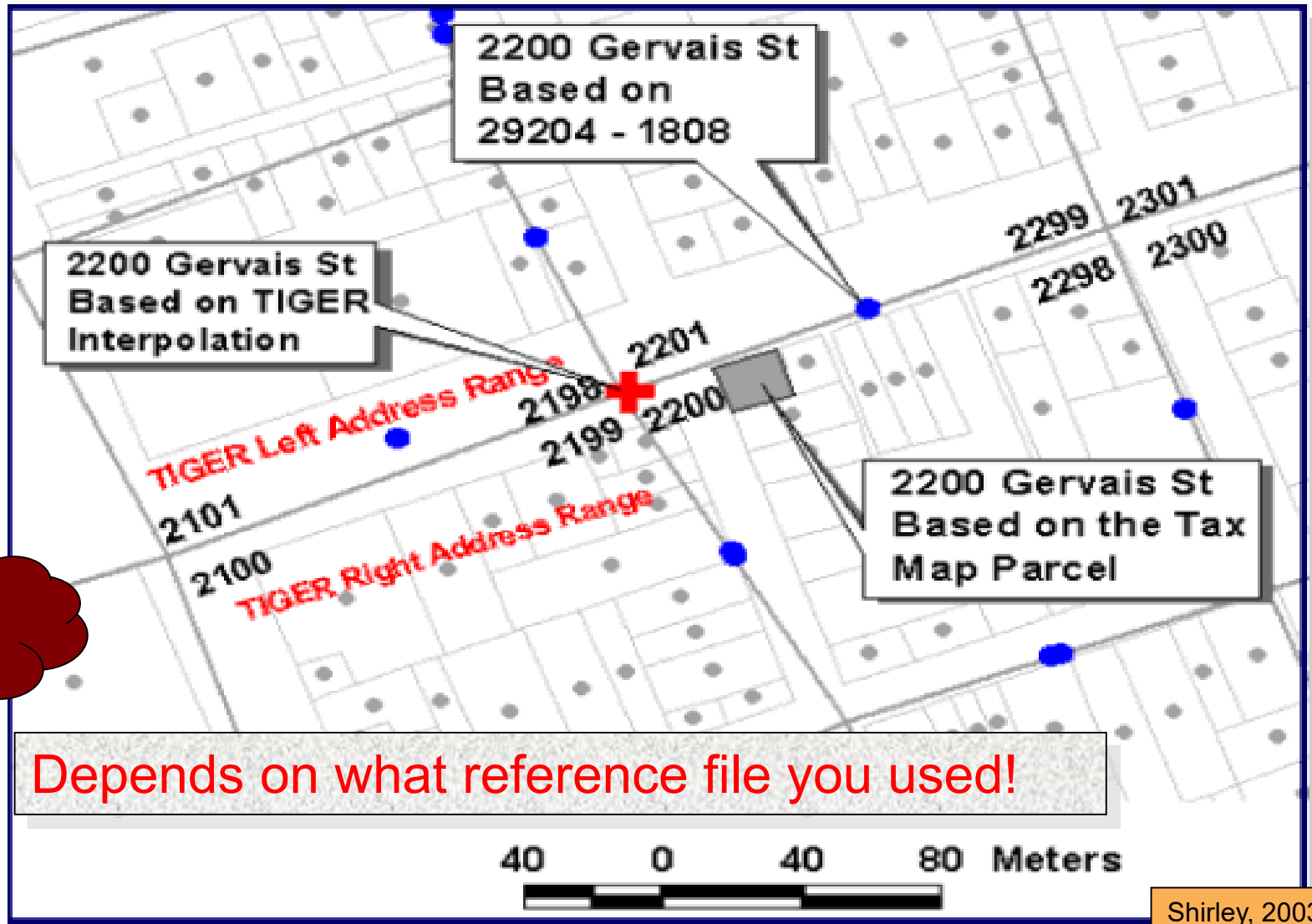Address difference: 1147 – 1101 = 46

Address range: 1199 – 1101 = 98

Address % along its street segment = 46 / 98 = 47%

Coordinate change in X = 6391 – 5000 = 1391

Interpolated **X = (0.47 * 1391) + 5000 = 5653.77**

Address Matching

Where is 2200 Gervais Street?

2200 Gervais St Based on 29204 - 1808

2200 Gervais St Based on TIGER Interpolation

2299   2301
2298   2300

TIGER Left Address Range
2198   2201
2199   2200
2101
2100
TIGER Right Address Range

2200 Gervais St Based on the Tax Map Parcel

Which is the best?

Depends on what reference file you used!

40   0   40   80   Meters

Shirley, 2003

Geocoding Platforms

Local	Online
Local geocoder ([city of Chicago Geostreets](#))	U.S. Census Geocoder
	Open Street Map Nominatim
Create geocoder based on street centerlines	ESRI World Geocoder
	OpenCage
	SAS Proc Geocode
ESRI street map premium	Google/Bing/Mapquest API

ESRI World Geocoder

- Pros:

  - Most accurate option

  - Fast processing

- Cons:

  - Expensive and difficult credit system

  - Higher learning curve than alternatives

Google/Bing/Mapquest API

- Pros:

  - Can use Stata. R, and other tools

  - Many are already be familiar with these services

- Cons:

  - Limited free queries per IP per day

  - Less transparent on quality of matches

  - Enterprise API key is expensive

U.S. Census Geocoder

- Pros:
  - Completely free
  - Easy to use
- Cons:
  - Limited to 1,000 addresses
  - Lower accuracy than alternatives

SAS Proc Geocode

- Pros:

  - Free (inside UI)

  - Handles unlimited addresses

- Cons:

  - Basic knowledge of SAS required
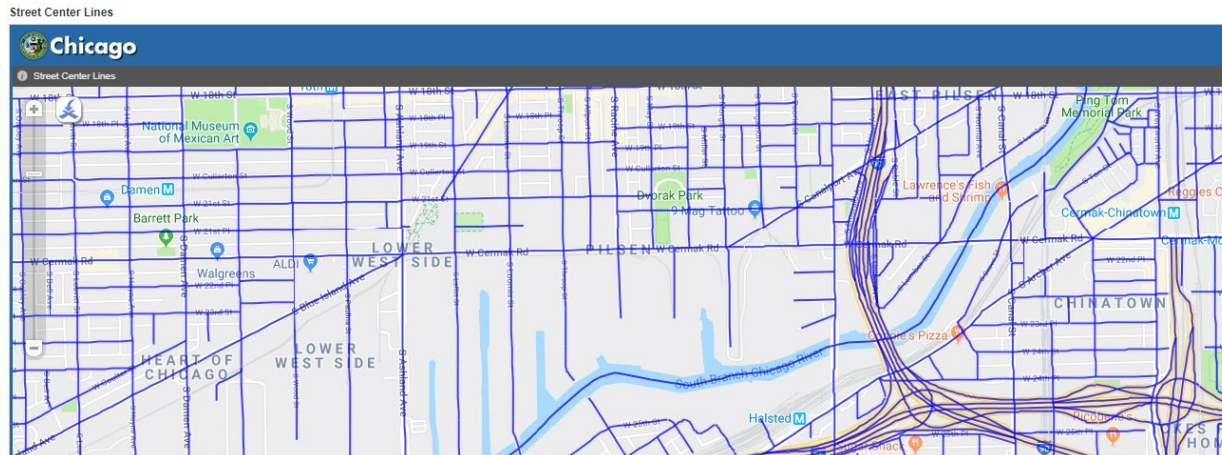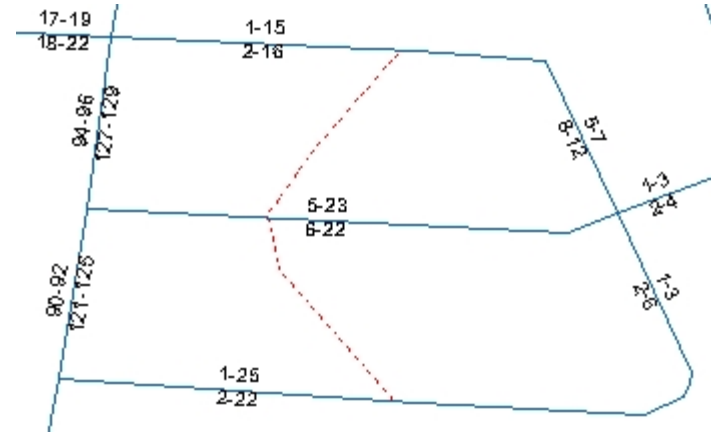
  - Lower accuracy than alternatives

Open Street Map Nominatim

- Pros:

  - Free

  - Across the globe coverage

  - Setup local server

- Cons:

  - Lower accuracy than alternatives

Local Geocoders

- Pros:

  - Highly accurate

  - Usually free to use

- Cons:

  - Limited to a single city

  - Need to create or learn for each city

Local Geocoders: StreetMap Premium

- Pros:

  - Highly accurate

  - One time cost

- Cons:

  - Expensive, need to purchase separate dataset for each state

StreetMap Premium

Enriched global data optimized for map display, geocoding, and routing. Ready-to-use on premises.

RCC-GIS Geocoding Service: Formatting Data for Processing

https://gis.rcc.uchicago.edu/content/rcc-gis-geocoding-service

Based on ESRI world Geocoder – direct link

Acceptable headers:

ID

ADDRESS

NEIGHBORHOOD

CITY

SUBREGION

REGION or STATE or ST

POSTAL or ZIP or ZIP CODE

COUNTRYCODE

ADDRESS	CITY	STATE	ZIP
1114 E 62ND ST	CHICAGO	IL	60637
12033 12TH AVE NE	SEATTLE	WA	98125
25500 SW 137TH AVE	HOMESTEAD	FL	33032
14765 CHAMPIONS VIEW PKWY	ALPHARETTA	GA	30004
4001 E MADISON ST	PHOENIX	AZ	85034

Longitude	Latitude	Match Score
-87.5987642	41.7823265	100
-122.315383	47.7167854	100
-80.4128246	25.5304734	100
-84.2738679	34.1423195	100
-111.995103	33.4462225	100

Generating Google Geocoder API Key

1.   Go to the **Google** Cloud Platform Console.

https://developers.google.com/maps/documentation/geocoding/get-api-key

1.   From the Project drop-down menu, select or **create** the project for which you want to add an **API key**.

2.   From the Navigation menu, select **APIs** & Services > Credentials.

3.   On the Credentials page, click **Create** credentials > **API key**. ...

4.   On the dialog, click Restrict **Key**

```
{
   "results" : [
      {
         "address_components" : [
            {
               "long_name" : "5801",
               "short_name" : "5801",
               "types" : [ "street_number" ]
            },
            {
               "long_name" : "South Ellis Avenue",
               "short_name" : "S Ellis Ave",
               "types" : [ "route" ]
            },
            {
               "long_name" : "Hyde Park",
               "short_name" : "Hyde Park",
               "types" : [ "neighborhood", "political" ]
            },
            {
               "long_name" : "Chicago",
               "short_name" : "Chicago",
               "types" : [ "locality", "political" ]
            },
            {
               "long_name" : "Chicago",
               "short_name" : "Chicago",
               "types" : [ "administrative_area_level_3", "political" ]
            },
            {
               "long_name" : "Cook County",
               "short_name" : "Cook County",
               "types" : [ "administrative_area_level_2", "political" ]
            },
            {
               "long_name" : "Illinois",
               "short_name" : "IL",
               "types" : [ "administrative_area_level_1", "political" ]
            },
            {
               "long_name" : "United States",
               "short_name" : "US",
               "types" : [ "country", "political" ]
            },
            {
               "long_name" : "60637",
               "short_name" : "60637",
               "types" : [ "postal_code" ]
            }
         ],
         "formatted_address" : "5801 S Ellis Ave, Chicago, IL 60637, USA",
         "geometry" : {
            "location" : {
               "lat" : 41.7886079,
               "lng" : -87.5987133
            },
            "location_type" : "ROOFTOP",
            "viewport" : {
               "northeast" : {
                  "lat" : 41.7899568802915,
                  "lng" : -87.5973643197085
               },
               "southwest" : {
                  "lat" : 41.7872589197085,
                  "lng" : -87.60006228029151
               }
            }
         },
         "place_id" : "ChIJRUhmExYpDogRyvXSAM9lf18",
         "plus_code" : {
```
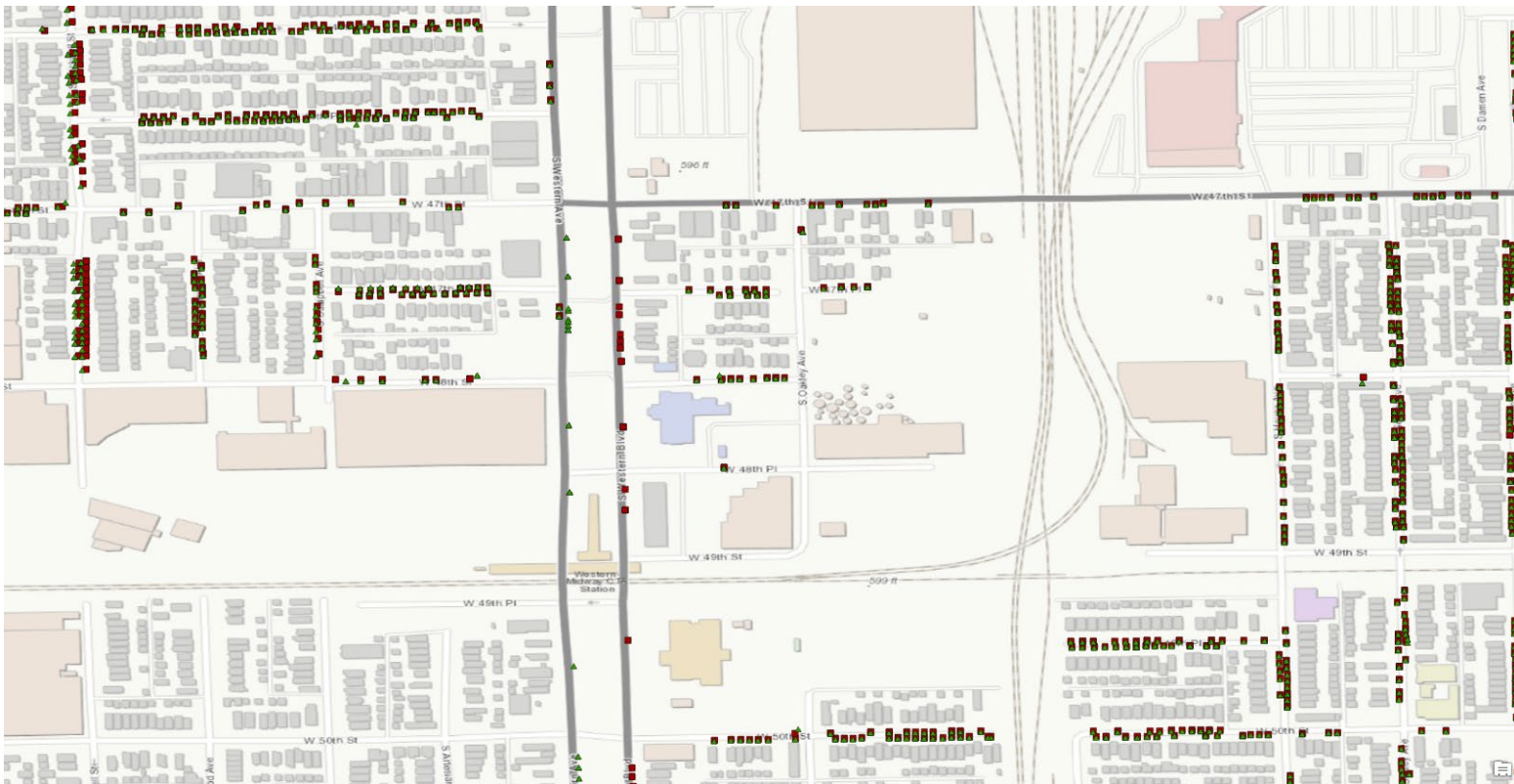
Geocoding Error

ESRI world locator could not locate W 43rd St.

Red rectangles - street map premium   **Green triangles** - ESRI world geocoder
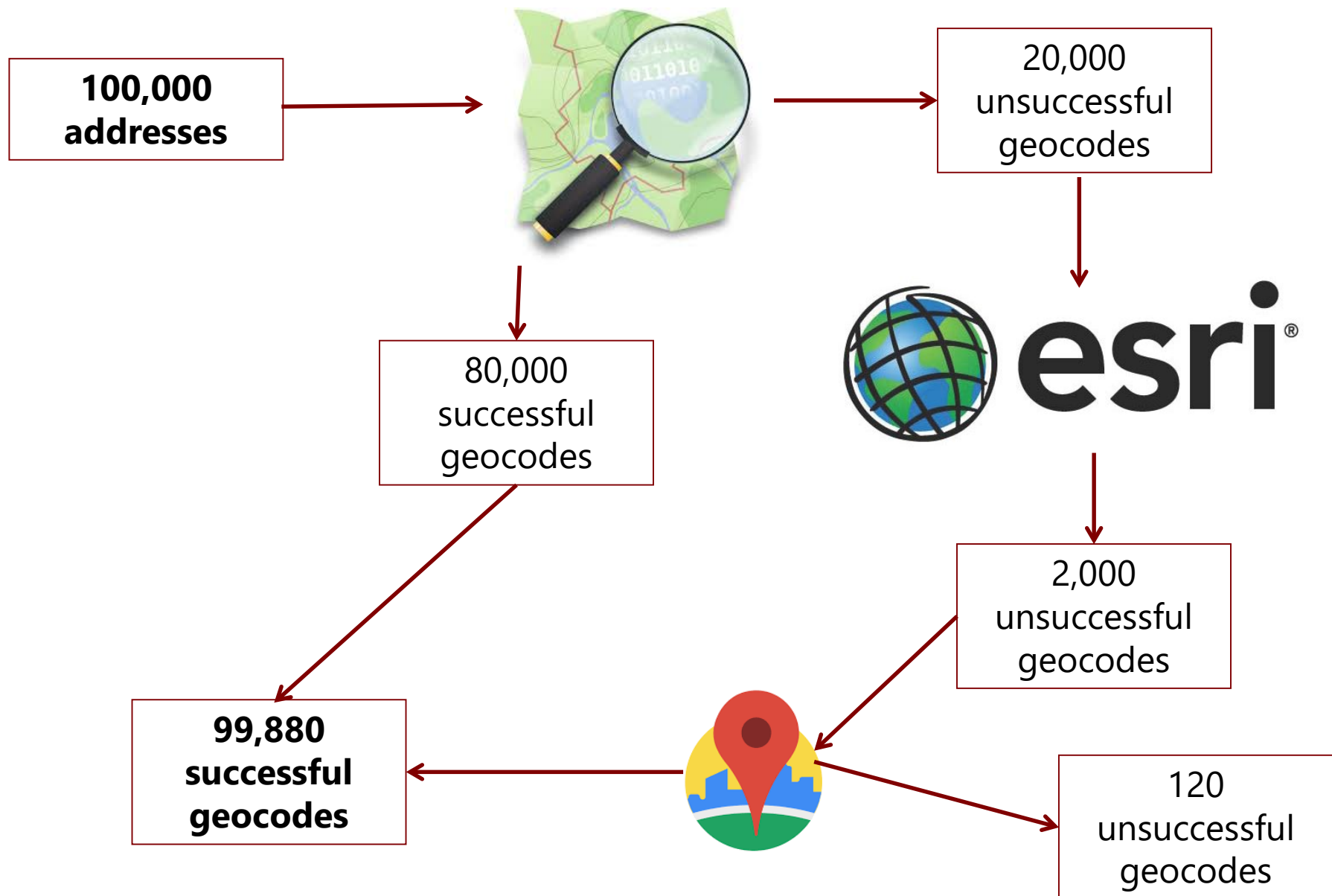
Geocoding Error

S Western Ave is one-way street (the green triangles for blvd are on the wrong side)

Red rectangles - street map premium   **Green triangles** - ESRI world geocoder

Hybrid Geocoding Approaches

100,000 addresses

20,000 unsuccessful geocodes

80,000 successful geocodes

2,000 unsuccessful geocodes

99,880 successful geocodes

120 unsuccessful geocodes

Hands on Training in R

1.Google Geocoder API

2.OSM Geocoder

3.Compare the distance between Geocoded locations

Questions?

gis-help[at]rcc.uchicago.edu