82 Below - Visualizing Lead Contamination Risk in Chicago

Team: Sebastian Buzenas, Obaid Bin-Mahfoudh, Gavin Coffer, Pranav Kuchibhotla, Thailer Simmons

Instructor: Robert Ellis

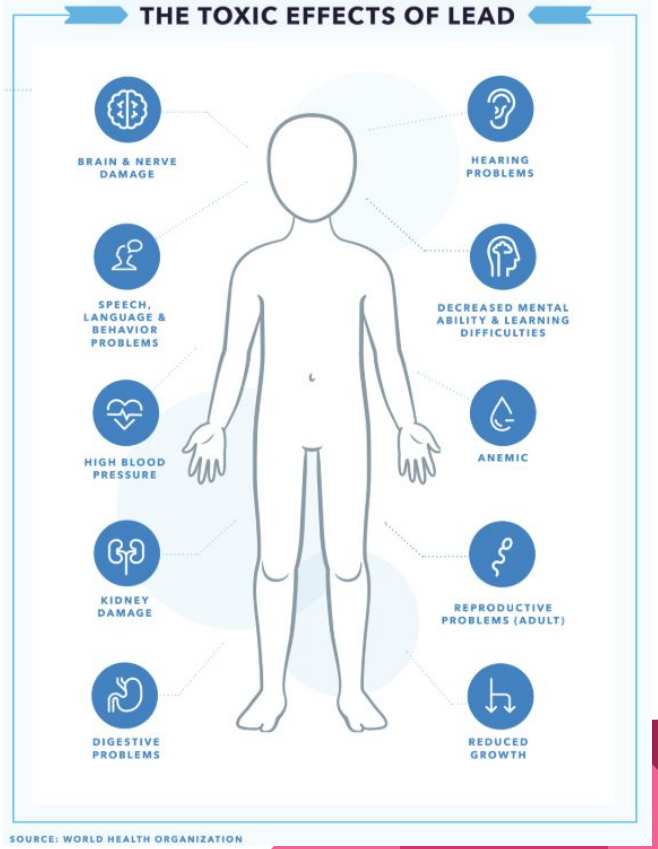
Subject Expert: Elin Betanzo, CEO of Safe Water Engineering LLC

Problem Statement

Chicago's outdated lead service lines pose a significant *public health risk*, especially to *children*.

Exposure to lead is linked to *cognitive and developmental issues*.

Our Goal: Create an accessible, interactive visualization of lead contamination risk using geospatial data as well as empower the public and policymakers to take action by providing a clear risk assessment tool.



Existing Work and Data

Lead Exposure Risk factor:

https://storymaps.arcgis.com/stories/87b9c4623fd04ad49793bfd414d3ae65

https://www.vox.com/a/lead-exposure-risk-map

The two links above display a lead exposure map of both Chicago and of the whole country.

https://sli.chicagowaterquality.org/

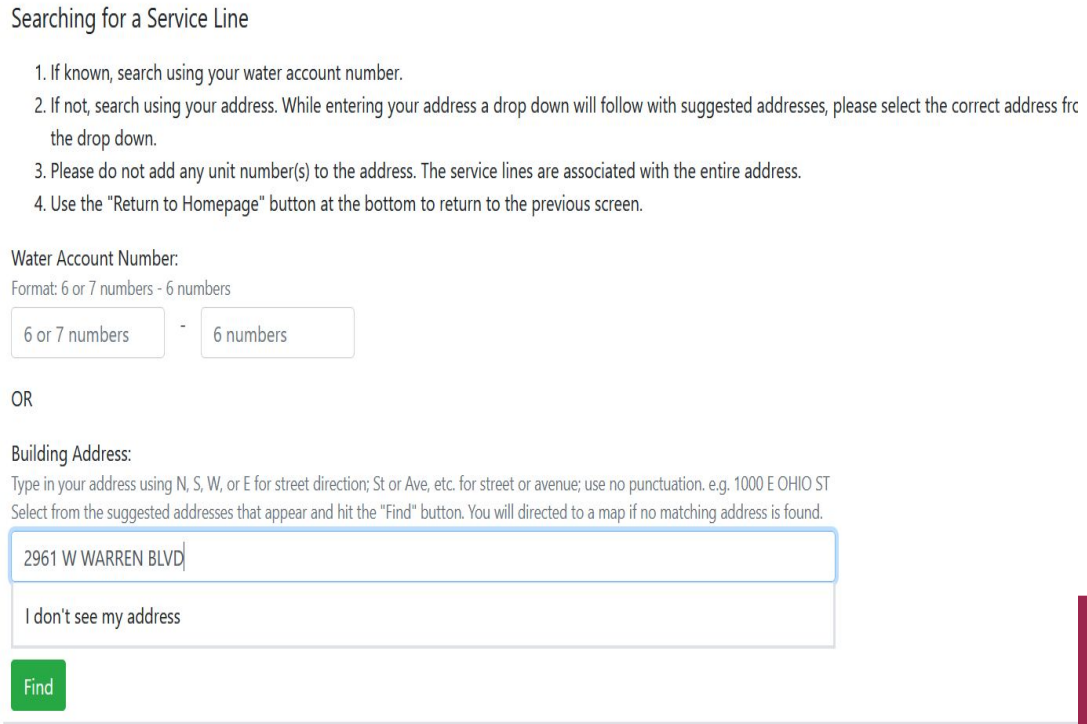
The link above is the cities website to look up service lines

The first uses Single Family or Two Flat Residences, Population without Health Insurance, People of Color (Black and/or Latinx), Presence of Children, neighborhood income and the age of the house as factors.

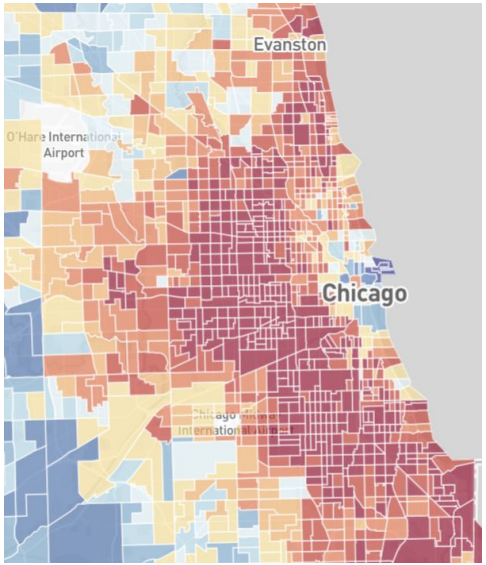
The second exposure risk/lead risk score is purely based off of neighborhood income and their date of development.

City’s website

When you can’t find an address you have to manually go through a map and find the service lines close to your address.

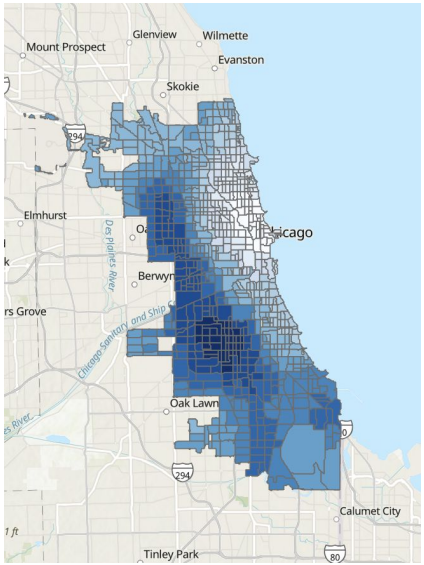


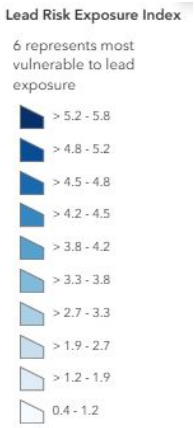
Existing Lead Exposure Maps



Lead exposure risk ranked from 1-10

Vox's exposure map of America





Sebastián Villa-Rodriguez's exposure map created from ArcGIS

Team Organization

- Team Structure: Collaborative, cross-functional approach
- Roles:
  - *Sebastian Buzenas*: Spokesperson
  - *Obaid Bin-Mahfoudh*: Leader
  - *Gavin Coffer*: Editor
  - *Pranav Kuchibhotla*: Data manager
  - *Thailer Simmons*: Recorder
- Tools Used: Github, OneDrive, and Discord for collaboration

Progress Toward Goals

Data acquisition: Geocoded Chicago residential addresses, Geocoded Chicago service line addresses, acquired coordinates for schools apart of the Chicago Public School district, acquired demographic data

Initial Visualizations: Initial visualization using ArcGIS

Website Development: 82below.vercel.app

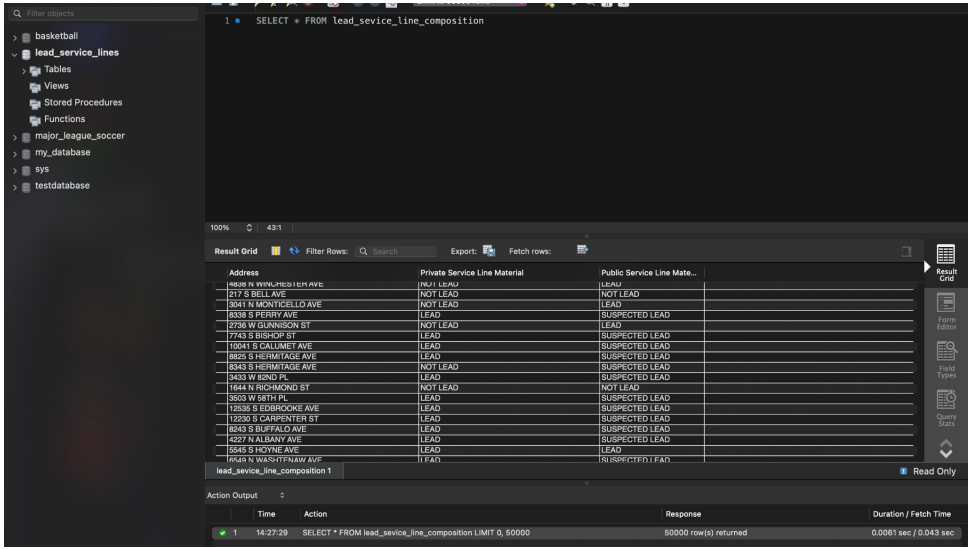
Database organization: MySQL workbench

Tools used: City of Chicago GIS Geocoder, Census geocoder, ArcGIS, SQL, Python

MySQL Database

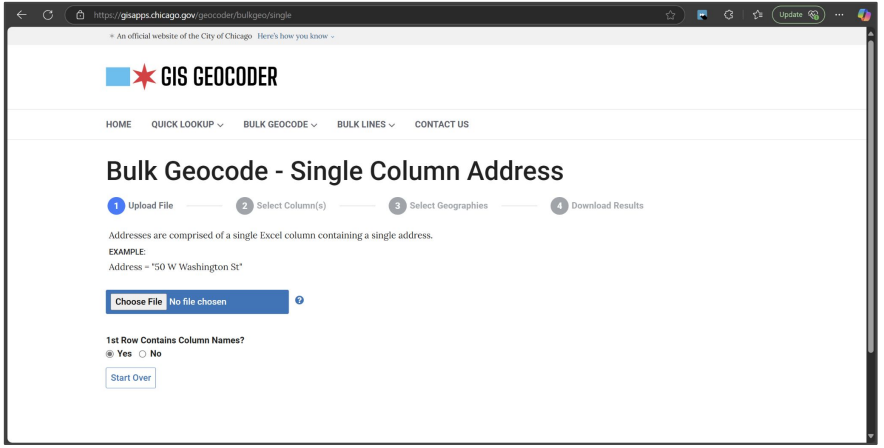
MySQL allows for better organization and manipulation of the datasets we are obtaining from multiple sources.

Combination of two tables can also be made easier with a MySQL database.



Geocoding

- Strategies:
  - Cook County Tax Assessor Address Joins
  - One-to-one Address Mapping
  - Chicago Geocoding Website
- Limitations:
  - Data quantity constraints of methods
  - Time required for each method
  - Accuracy of geocoders based on data quality
  - Inconsistencies between methods



Visualisation

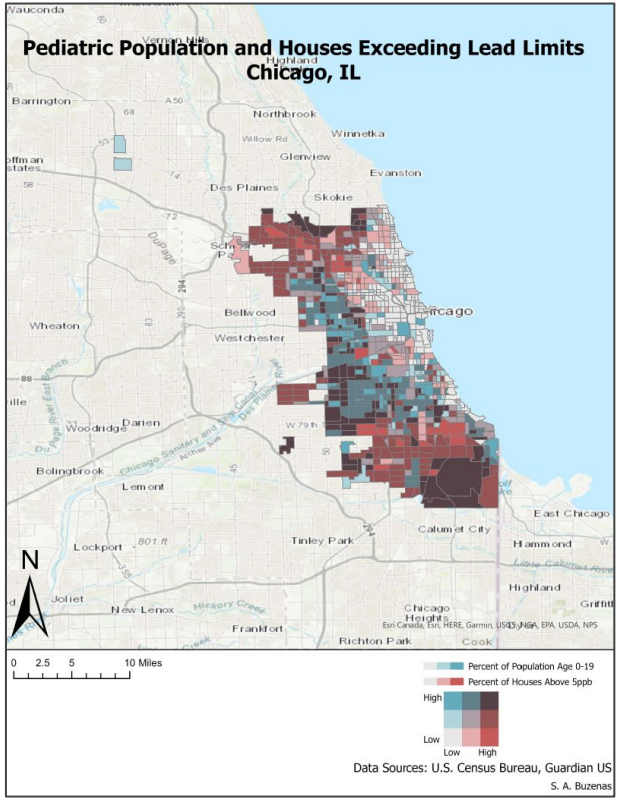
- Lead Risk Factors to consider
  - Youth Population
  - Median Income
  - Health Insurance Coverage

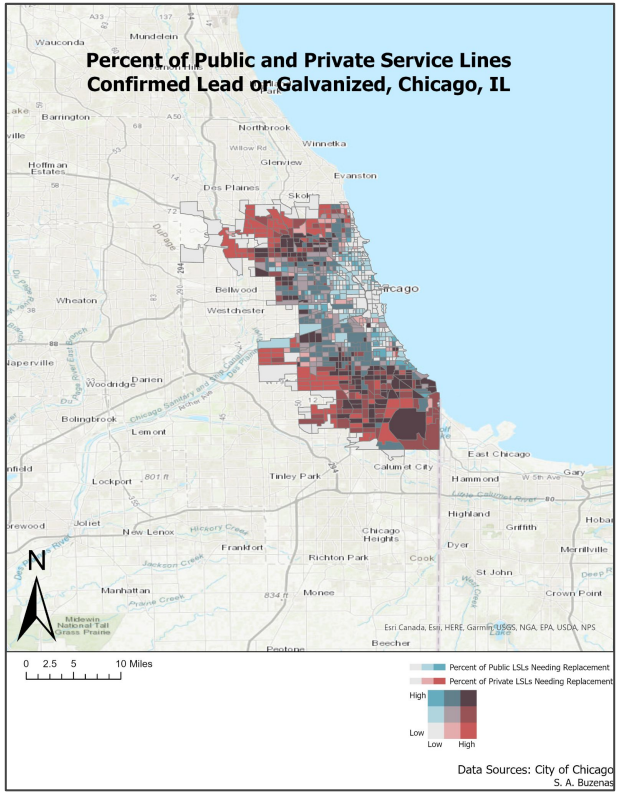
- Public and Private SL Composition
  - Geocoding Challenges
  - Hotspots

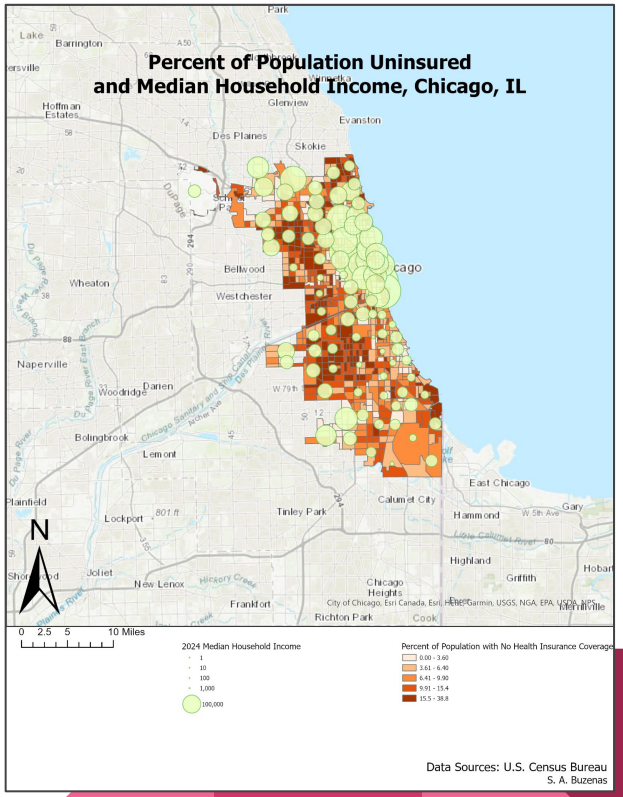
- Lead and Copper Rule (Guardian Data):
  - Lead exposure not to exceed 15ppb in drinking water
  - Recommendation for schools/childcare facilities not to exceed 5ppb (not enforced)

ArcGIS

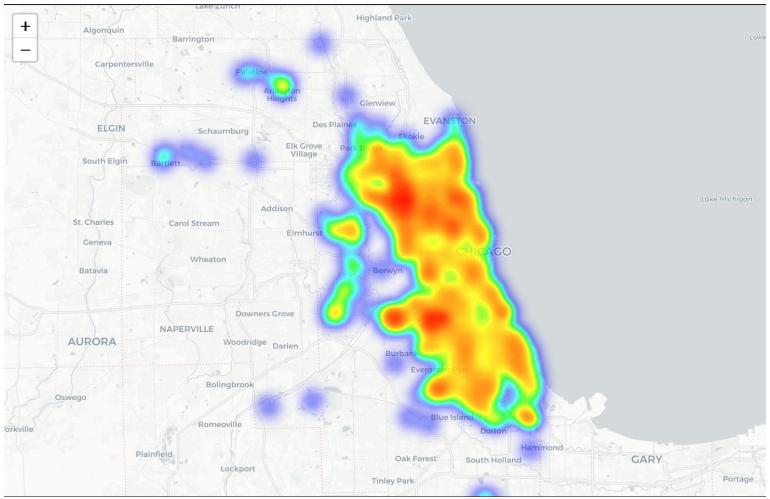
Visualisation

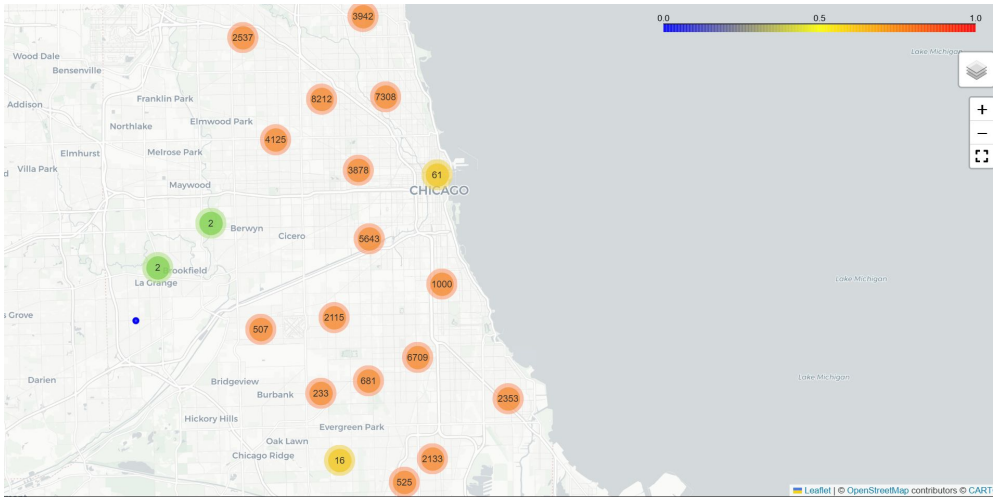






Visualisation



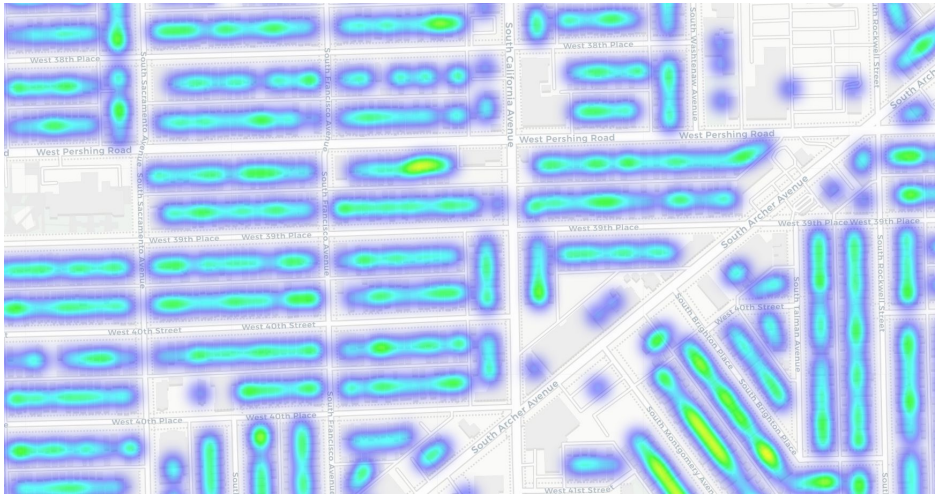


We used data science techniques like Kernel Density Estimation (KDE) and geospatial mapping to identify high-risk areas.

Lead Density Visualisation

To measure the concentration of lead service lines, we used KDE. We computed density scores using Gaussian KDE, which smooths data points to reveal broader trends rather than just individual locations.

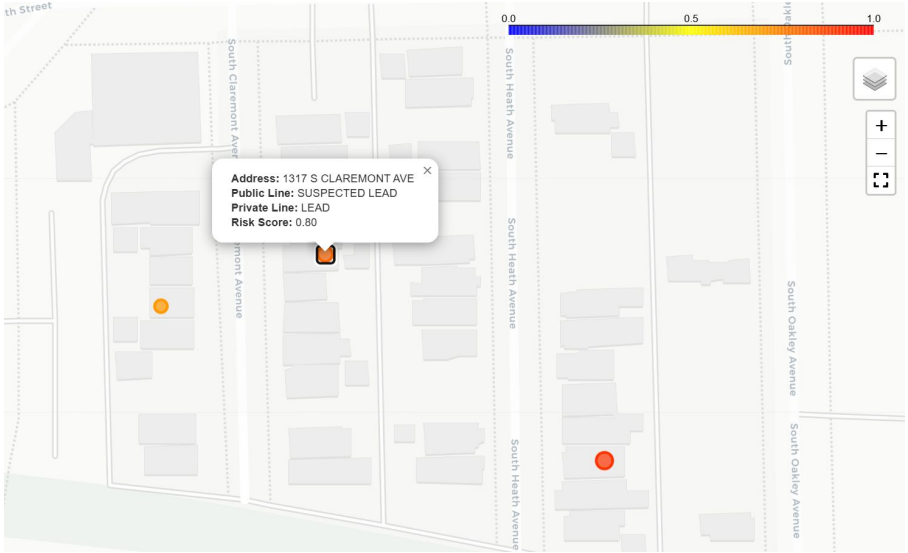
Each data points intensity is determined by its KDE density score, helping visualize problem areas more effectively.



Lead Risk Visualisation

The color scale transitions from blue (low risk) to red (high risk) making it easy to identify critical areas.

The marker cluster groups nearby properties, improves readability and clicking on a property reveals details like its address, service line materials, and the risk score.



Data Exploration

Service Line: We have a total of 487,347 data points

We have around 3,979 duplicate addresses for a total of 483,368 unique addresses

Assessor: We have a total of 884,116 data points

What causes duplicates?

Data Exploration - Service Line Duplicates

Some are straight duplicates

7110 S PARNELL AVE	NOT LEAD	LEAD
7110 S PARNELL AVE	NOT LEAD	LEAD

Some are duplicates with one saying not lead and the other saying Galvanized

Address	Private Service Line Materia	Public Service Line Material
3250 W ADAMS ST	NOT LEAD	GALVANIZED REQUIRING REPLACEMENT
3250 W ADAMS ST	NOT LEAD	NOT LEAD

Some addresses duplicates have different values. We can conclude that these are different service lines

Address	Private Service Line Materia	Public Service Line Material
2301 S LAKE SHORE DR	SUSPECTED LEAD	SUSPECTED LEAD
2301 S LAKE SHORE DR	SUSPECTED LEAD	SUSPECTED LEAD
2301 S LAKE SHORE DR	SUSPECTED LEAD	SUSPECTED LEAD
2301 S LAKE SHORE DR	SUSPECTED LEAD	SUSPECTED LEAD
2301 S LAKE SHORE DR	SUSPECTED LEAD	SUSPECTED LEAD
2301 S LAKE SHORE DR	SUSPECTED LEAD	SUSPECTED LEAD
2301 S LAKE SHORE DR	LEAD	SUSPECTED LEAD
2301 S LAKE SHORE DR	NOT LEAD	NOT LEAD
2301 S LAKE SHORE DR	NOT LEAD	NOT LEAD
2301 S LAKE SHORE DR	NOT LEAD	NOT LEAD

Data Exploration- Assessor Duplicates

Duplicates on address due to multiple apartment numbers.

pin	address	propert	propert	propert	mailing_address
1.32E+13	6040 W EASTWOOD AVE	301	CHICAGO	60630-317	6040 W EASTWOOD AVE
1.32E+13	6040 W EASTWOOD AVE	202	CHICAGO	60630-317	6040 W EASTWOOD AV#202
1.32E+13	6040 W EASTWOOD AVE	303	CHICAGO	60630-317	6040 W EASTWOOD #303
1.32E+13	6040 W EASTWOOD AVE	101	CHICAGO	60630-317	6040 W EASTWOOD AV
1.32E+13	6040 W EASTWOOD AVE	302	CHICAGO	60630-317	6040 W EASTWOOD AVE302
1.32E+13	6040 W EASTWOOD AVE	203	CHICAGO	60630-317	6040 W EASTWOOD AVE203
1.32E+13	6040 W EASTWOOD AVE	201	CHICAGO	60630-317	6040 W EASTWOOD AVE

Some addresses have different latitude and longitudes for same address.

Data Exploration- Expanding Addresses

Some service line addresses are a range of addresses

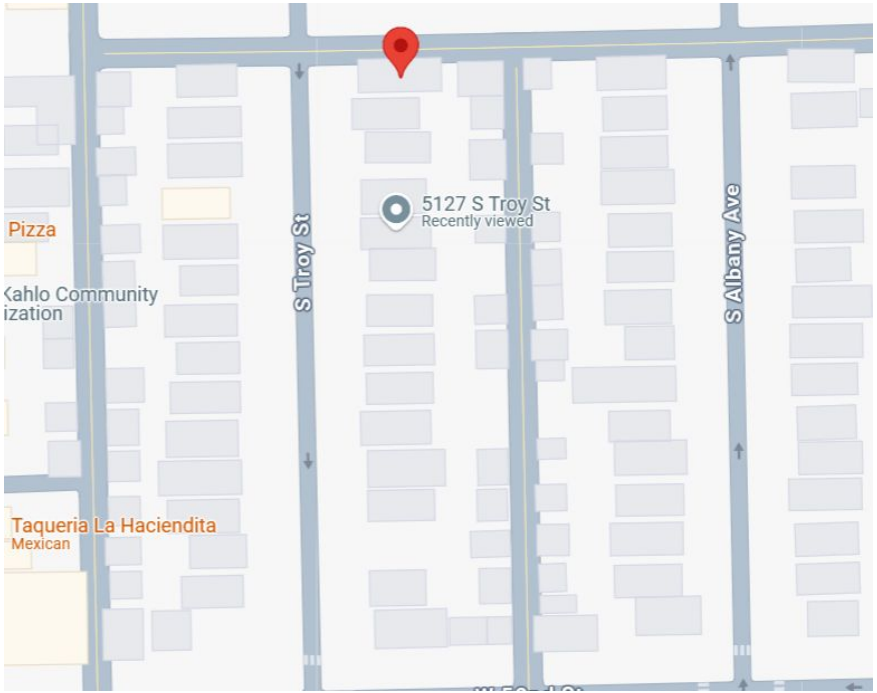
There are 30,492 expandable addresses. Of those 50 are ending in 00. Of those 4 start with a number.

After partial expanding we get 573,762 addresses. An increase of 86,415 addresses.

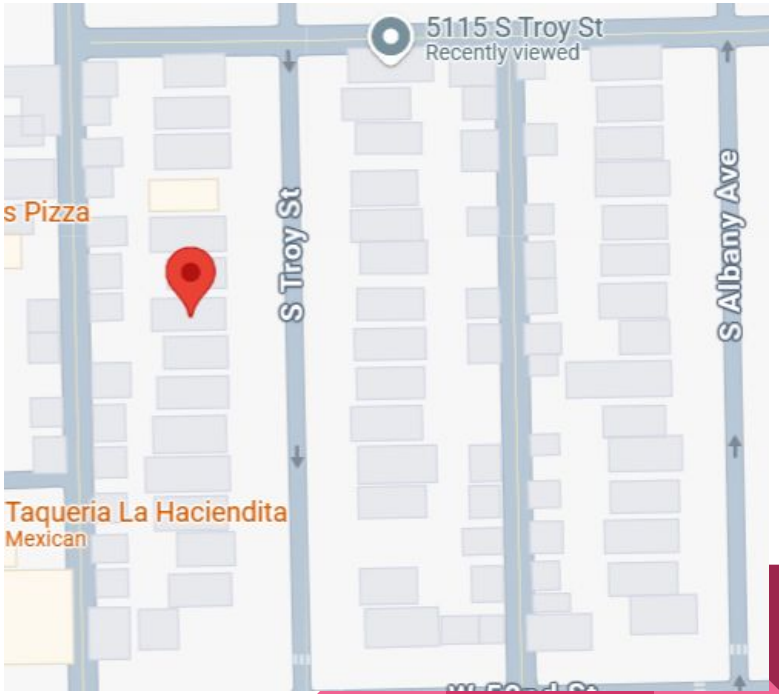
This is where it gets kind of fuzzy on how to maintain service line data. We can expand for merges but the unique number of service lines does not increase

Address	Private Service Line Materia	Public Service Line Material
6355 57 S WASHTENAW AVE	SUSPECTED LEAD	SUSPECTED LEAD
63 65 E 79TH ST	SUSPECTED LEAD	SUSPECTED LEAD
5127 00 S TROY ST	SUSPECTED LEAD	LEAD
5115 00 S TROY ST	SUSPECTED LEAD	LEAD
5132 00 S TROY ST	SUSPECTED LEAD	LEAD
4514 00 N CALIFORNIA AVE	NOT LEAD	NOT LEAD
2300 00 N CENTRAL AVE	NOT LEAD	NOT LEAD
671 75 N MICHIGAN AVE	GALVANIZED REQUIRING REP	GALVANIZED REQUIRING REPLACEMENT
671 75 N MICHIGAN AVE	GALVANIZED REQUIRING REP	GALVANIZED REQUIRING REPLACEMENT

5115 vs 5127 S Troy St



5132 vs 5115 S Troy



Data Exploration - Merging addresses

Merge on address and mailing address, have to make sure we do not duplicate matches

matched_a	match_typ	property_c	property_z	latitude	longitude	township_	school_ele	school_hs_	Private Ser	Public Service Line Material
1215 W CC	address	CHICAGO	60657-142	41.945045	-87.6597	Lake View	HAMILTON	LAKE VIEW	NOT LEAD	LEAD
4838 N WI	address	CHICAGO	60640-400	41.969960	-87.6782	Lake View	MCPHERS	AMUNDSE	NOT LEAD	LEAD
217 S BELL	address	CHICAGO	60612-297	41.878085	-87.6824	West Chic	DETT	MARSHALL	NOT LEAD	NOT LEAD
3041 N MC	address	CHICAGO	60618-661	41.93678	-87.7182	Jefferson	REILLY	SCHURZ H	NOT LEAD	LEAD
8338 S PEF	address	CHICAGO	60620-123	41.742430	-87.6272	Lake	WESTCOT	HIRSCH H	LEAD	SUSPECTED LEAD
2736 W GL	address	CHICAGO	60625-280	41.969943	-87.6978	Jefferson	BUDLONG	AMUNDSE	NOT LEAD	LEAD
7743 S BIS	address	CHICAGO	60620-412	41.752766	-87.6598	Lake	COOK	HYDE PAR	LEAD	SUSPECTED LEAD
10041 S CA	address	CHICAGO	60628-212	41.711567	-87.6148	Hyde Park	BENNETT	HARLAN H	LEAD	SUSPECTED LEAD
8825 S HEI	address	CHICAGO	60620-494	41.7332	-87.6663	Lake	FOSTER PA	HARLAN H	LEAD	SUSPECTED LEAD
8343 S HEI	address	CHICAGO	60620-463	41.741764	-87.6665	Lake	FOSTER PA	HIRSCH H	NOT LEAD	SUSPECTED LEAD

When finding matches we have 453,381 matched on address and 5,389 matches on mailing address. We have a total of 114,992 addresses with no direct match.

This is compared to 417,089 matched on address and 4044 matched on mailing address before expanding. Here we have a total of 66,214 addresses without a match.

Merge on Geocode

Using a cKDTree and based off of euclidean distance used the geocoded data from census to connect assessor addresses to service line addresses.

Unique ID	Address	Match_Sta	Match_Typ	Matched_A	Geodata	Lion_Code	Lion	Private Ser	Public Ser	Lat	Long	Connectec	Num_Connected_Addresses
330542	6620 N OA	Match	Non_Exact	6620 S OAI	-87.79090	6.55E+08	R	SUSPECTE	SUSPECTE	41.77227	-87.7909	['7000 W 7	191
501887	6825 N HA	Match	Non_Exact	6825 S HAI	-87.80070	1.12E+08		NOT LEAD	NOT LEAD	41.77324	-87.8007	['7113 W 7	150
305278	5001 S LOI	Match	Non_Exact	5001 S LOI	-87.75760	6.17E+08	L	SUSPECTE	SUSPECTE	41.80188	-87.7576	['4852 S LC	96
188611	4801 N CE	Match	Non_Exact	4801 S CEI	-87.76262	6.43E+08	L	SUSPECTE	SUSPECTE	41.80543	-87.7626	['4901 S LU	89
510604	12001 S TC	Match	Exact	12001 S TC	-87.55923	1.12E+08	L	NOT LEAD	NOT LEAD	41.67886	-87.5592	['12035 S T	76
100345	1833 N LO	Match	Exact	1833 N LO	-87.76068	6.18E+08	R	SUSPECTE	SUSPECTE	41.91504	-87.7607	['1817 N LC	68
305238	5330 W 47	Match	Exact	5330 W 47	-87.75577	1.12E+08	R	SUSPECTE	SUSPECTE	41.80751	-87.7558	['4812 S LC	67
471203	W JACKSO	Match	Non_Exact	JACKSON I	-88.083853609385,41.95068415			NOT LEAD	NOT LEAD	41.95068	-88.0839	['1720 W R	62
569900	1000 N KIN	Match	Exact	1000 N KIN	-87.64410	6.42E+08	L	NOT LEAD	NOT LEAD	41.89879	-87.6441	['1000 N KI	62
596235	2642 N HE	Match	Exact	2642 N HE	-87.67211	6.42E+08	L	NOT LEAD	NOT LEAD	41.92964	-87.6721	['2705 N HI	57
388725	2700 W ME	Match	Exact	2700 W ME	-87.69533	6.05E+08	R	LEAD	SUSPECTE	41.94032	-87.6953	['2649 W M	55
305266	5052 S LOT	Match	Non_Exact	5052 S LOT	-87.75895	1.12E+08	R	SUSPECTE	SUSPECTE	41.80095	-87.759	['4946 S LIN	54
548580	1309 S WC	Match	Exact	1309 S WC	-87.67119	6.05E+08	L	NOT LEAD	NOT LEAD	41.86402	-87.6712	['1810 W H	50
577422	2712 1/2 W	Match	Non_Exact	2712 1/2 W	-87.69588	6.42E+08	R	NOT LEAD	NOT LEAD	41.95404	-87.6959	['3953 N W	49
549036	1412 S BLL	Match	Exact	1412 S BLL	-87.65670	6.05E+08	R	NOT LEAD	NOT LEAD	41.86319	-87.6567	['1407 W 1	47
305240	5401 W 47	Match	Exact	5401 W 47	-87.75790	1.12E+08	R	SUSPECTE	SUSPECTE	41.80737	-87.7579	['4831 S LC	44
516856	4532 W BE	Match	Exact	4532 W BE	-87.74144	6.05E+08	R	NOT LEAD	NOT LEAD	41.95716	-87.7414	['4216 N KC	44
178556	1324 S LOC	Match	Exact	1324 S LOC	-87.66158	6.46E+08	R	SUSPECTE	SUSPECTE	41.8649	-87.6616	['1428 W 1	43
518737	810 W 14T	Match	Exact	810 W 14T	-87.64725	6.53E+08	R	NOT LEAD	NOT LEAD	41.8624	-87.6473	['1421 S CA	42
178369	1831 W WA	Match	Exact	1831 W WA	-87.67283	6.38E+08	L	SUSPECTE	SUSPECTE	41.86583	-87.6728	['1844 W 1	41
515180	501 N CLIN	Match	Non_Exact	501 N CLIN	-87.64228	6.42E+08	R	NOT LEAD	NOT LEAD	41.89074	-87.6423	['501 N CLI	39
178373	2227 W WA	Match	Exact	2227 W WA	-87.68180	1.12E+08	L	SUSPECTE	SUSPECTE	41.86569	-87.6818	['2236 W 1	38
305242	5448 W 47	Match	Exact	5448 W 47	-87.75888	1.12E+08	L	SUSPECTE	SUSPECTE	41.80748	-87.7589	['4824 S LC	38
471761	12600 S TC	Match	Exact	12600 S TC	-87.55894	6.4E+08	R	NOT LEAD	NOT LEAD	41.66612	-87.5589	['12509 S T	38
569899	1239 N SEI	Match	Exact	1239 N SEI	-87.63830	6.18E+08	R	NOT LEAD	NOT LEAD	41.90449	-87.6383	['1235 N OI	37
481753	6615 N KEI	Match	Exact	6615 N KEI	-87.70929	6.18E+08	R	NOT LEAD	NOT LEAD	42.00175	-87.7093	['3130 W W	36
529275	W 51ST ST	Match	Exact	W 51ST ST	-87.761298219246,41.80006916			NOT LEAD	NOT LEAD	41.80007	-87.7613	['5006 S LU	36

Current Challenges and Solutions

Missing Assessor Data:

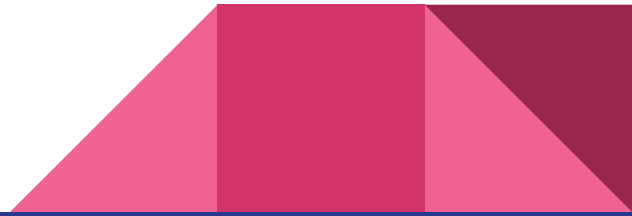
- Some addresses do not match between the Assessor dataset and Service Line dataset.
- Assessor dataset does not include all addresses
- **Solution**: Possibly geocode

Different Latitude/Longitude Coordinates for the Same Address:

- When geocoding the addresses we are getting some different longitude and latitudes
- **Solution:** Look at a reputable website like google and cross reference to find the more accurate source

Incomplete Geospatial Data:

- Some properties lack service line records.
- **Solution**: Add missing data using neighboring properties.



Future Work

- Once all of our data is compiled and matched, implementing an interactive visualization into the website may lead to some challenges.
- How can we further the website's ease of interaction and navigation, as well as displaying an *impactful* message about the importance of understanding your risk of lead exposure.
- Broadcasting the website and making it accessible to the public.

Thank You!