Assessing Chicago Lead Risk and Replacement Prioritization

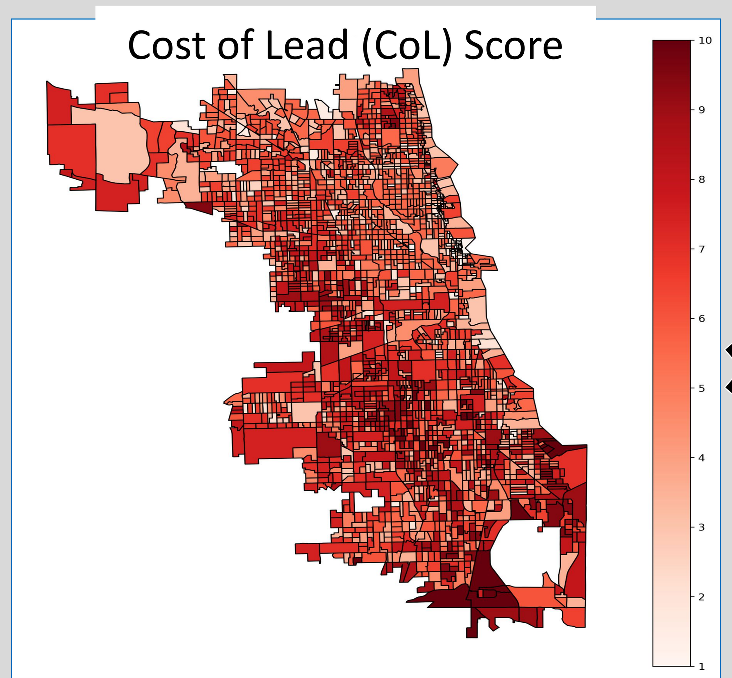
Obaid Bin-Mahfoudh, Sebastian Buzenas

Advisor: Robert Ellis (College of Computing, Illinois Institute of Technology), Subject Matter Expert: Elin Betanzo, P. E. (Safe Water Engineering)

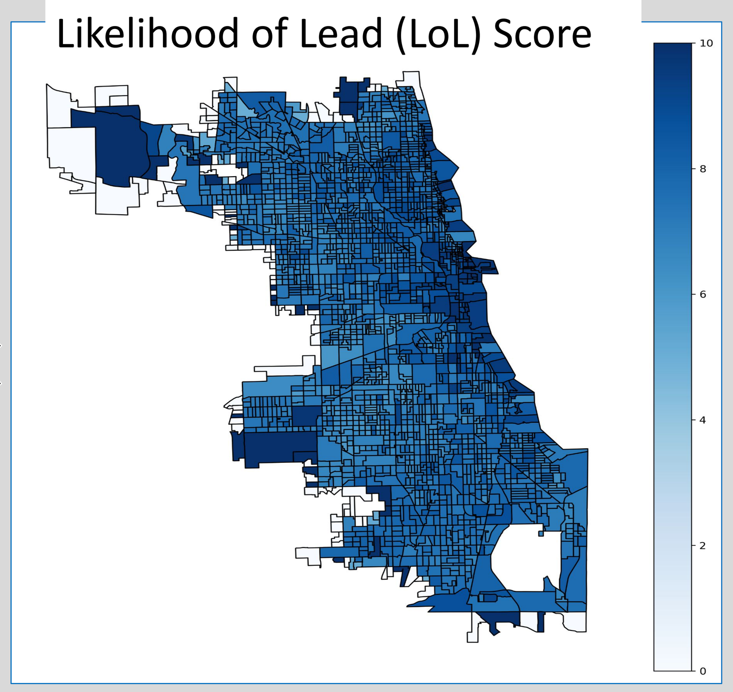
ABSTRACT

Chicago is home to the most lead service lines of any city in the United States. Lead metal is a toxic metal, known to cause reproduction and developmental harm to those exposed to any quantity of lead. Service lines are responsible for connecting homes and businesses to the city's water mains, which bring water to these buildings from treatment facilities. Service lines made from lead can leach into the drinking water and expose individuals to lead poisoning over time. While safety measures such as chemically lining the pipe interiors and encouraging residents to utilize lead filters on their taps, these are not permanent solutions, and Chicago must work to identify the communities that are more in need of government funding, as this money is limited and must be spent in the most effective way possible. Census demographics were used to establish a lead risk score to visualize what specific regions of Chicago may be more at risk for lead exposure, based on their access to resources and presence of vulnerable populations, namely children. Set to replace all of its lead lines in the next 2 decades, Chicago must work to identify which areas to focus on first to facilitate efficient replacement. Areas with lower income and less access to medical treatment are less able to respond to the threat that lead poses, and as such these communities must be addressed first when determining what lead service lines to replace first. Presence of children/adolescents also raises concern in regions in the city with higher lead risk due to lead's effect on child development; children exposed to lead at a younger age will suffer from lead poisoning for the rest of their lives.

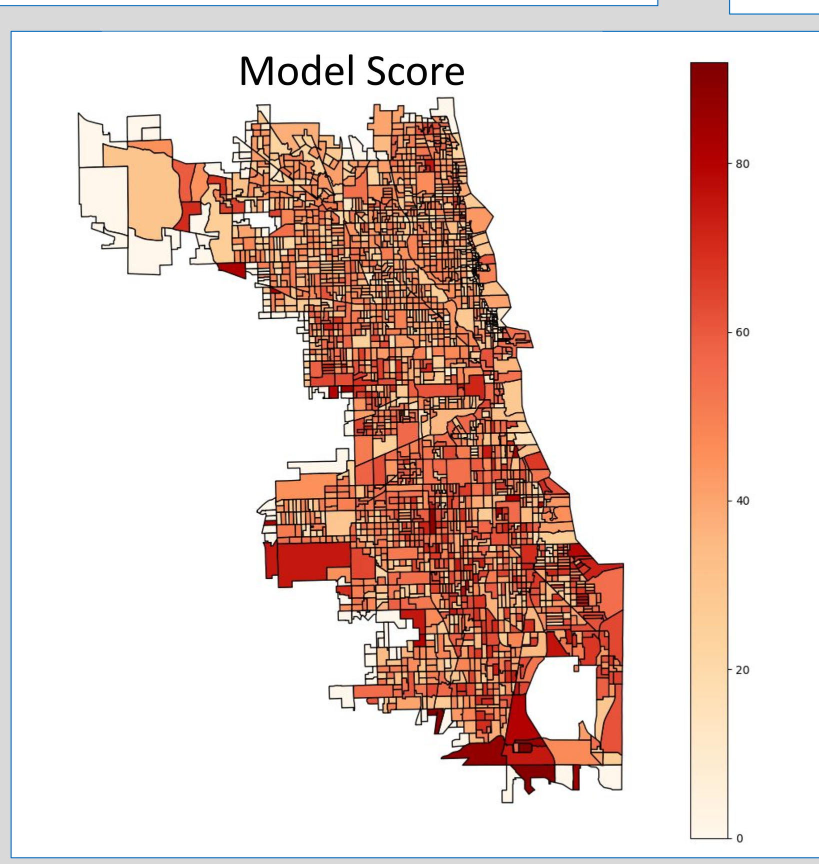
PRIORITIZATION MODEL



×



=



Prioritization map on census block groups.

CoL Score
CoL = 0.5 * ADI Score + 0.5 * Children under 5 score

Children under 5 score
50+ children: 10
11-50 children: 5
1-10 children: 3
No children: 1

LoL score
Lead and galvanized lines: 10.
Unknowns: 5
Non-Lead: 0

Averaged per block group

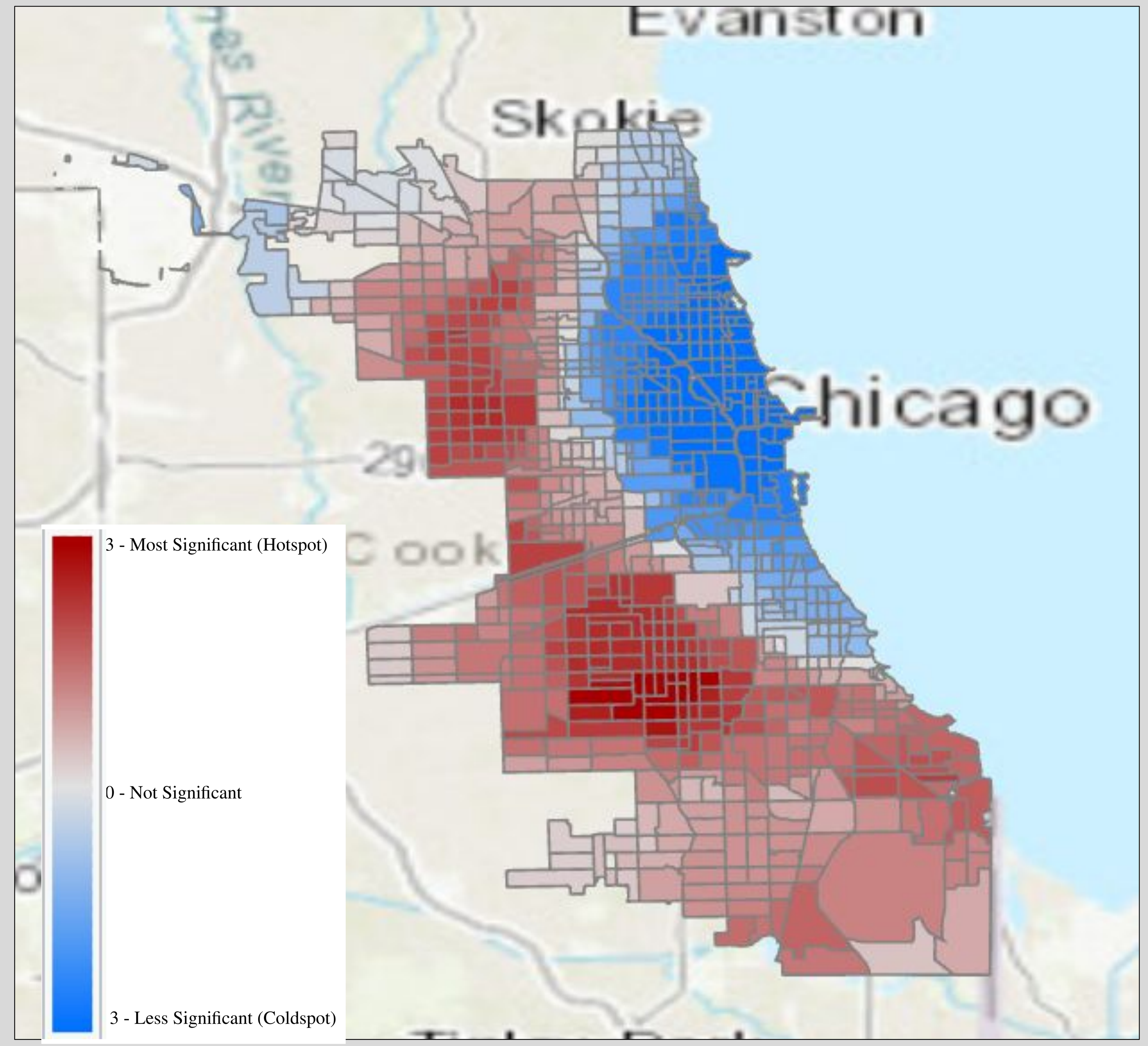
Model score = CoL * LoL

ETHIC CONCERN / STAKEHOLDERS

Lead exposure is not just a technical problem, its a public health and justice issue.

Children under five face the highest risks, especially in communities with high poverty or social vulnerability. Our research prioritizes areas where both the likelihood of lead and potential for harm are high, using factors such as the Area Deprivation Index and census demographics regarding income, access to medical care, and child population data.

- **Lead is a major health concern. If actions are not taken swiftly to replace service lines, the entire population of Chicago risks exposure for the next several decades.**



Combined Getis-Ord Gi* Bin Plot of Percent of Private/Public Lead Lines, Percent of Private/Public Suspected Lines, Percent of Youth Population, Percent of Population Uninsured by Medical Coverage, and Inverse Median Household Income

CONCLUSIONS

- According to the combined hotspot analysis, regions of Central and Northern Chicago are coldspots, indicating lower risk of lead exposure. Based on the service data, there is a low concentration of lead lines. Based on the demographic data, the residents of these cold regions have more resources and less vulnerable children in their population, so they are more able to respond to a lead threat if one were present.
- West And Southern Chicago, namely around the Englewood and Austin neighborhoods, are identified as hotspots, indicating either high population of lead service lines, lack of resources and presence of vulnerable populations, or both.
- Our risk model, which takes into account the social cost of lead and the likelihood of lead, effectively identifies block groups where lead service line replacement will have the greatest public health impact.

FUTURE CONSIDERATIONS

- As Chicago continues with their service line replacement efforts, the lead service inventory must be continuously revisited to track which communities are being freed of lead poisoning. Temporospatial analysis could be done to demonstrate change in service line composition across the city given enough time intervals to see if Chicago is properly prioritizing replacement.

REFERENCES

1. Esri. *How Hot Spot Analysis (Getis-Ord Gi) Works*. ArcGIS Pro, https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm. Accessed 24 Apr. 2025.
2. "Hot Spot Analysis (Getis-Ord Gi*)." *GIS&T Body of Knowledge*, UCGIS, https://gistbok-ltb.ucgis.org/page/27/concept/8344. Accessed 24 Apr. 2025.
3. City of Chicago. *Lead in Drinking Water*. Lead Safe Chicago, https://www.leadsafechicago.org/water#:~:text=Any%20Chicago%20resident%20concerned%20about%20lead%20in%20their,are%20done%2C%20schedule%20a%20pick-up%20by%20city%20representatives. Accessed 24 Apr. 2025.
4. Neighborhood Atlas - Log In. (2019). Wisc.edu. https://www.neighborhoodatlas.medicine.wisc.edu/download
5. States, U. (2025). Explore Census Data. Census.gov.
6. https://data.census.gov/table/ACSDT5Y2022.B01001?q=B01001&g=050XX00US17031
7. Safe Water Engineering. (2022, October 27). Safe Water Engineering. Safe Water Engineering. https://www.safewaterengineering.com/washington-dc

METHODOLOGICAL APPROACH

ArcGIS Pro Map Visualization

- Demographics used to determine risk scores for each census tract were obtained from the US Census Bureau's open data portal[6] (% youth population, % with no health insurance, median household income). Data used concerning the composition of the lead service lines, both public and private, were obtained directly from the City of Chicago through a FOIA request.

Getis-Ord Gi* Statistic

- The Gi* statistic is calculated to determine where features of high and low values are significantly grouped together. Gi* computes a local mean and standard deviation for each feature (census tracts in this case) by comparing the target feature value with a predetermined amount of neighboring features. A new mean is calculated for the neighborhood formed and is compared with the global mean to determine whether it is significantly above or below the global mean. This results in hotspots that take spatial relationships between tracts into account. The hotspot map shown is an amalgamation of seven separate hotspot maps, one for each factor used to determine the lead risk among the tracts.

Determining Neighbors for Spatial Analysis

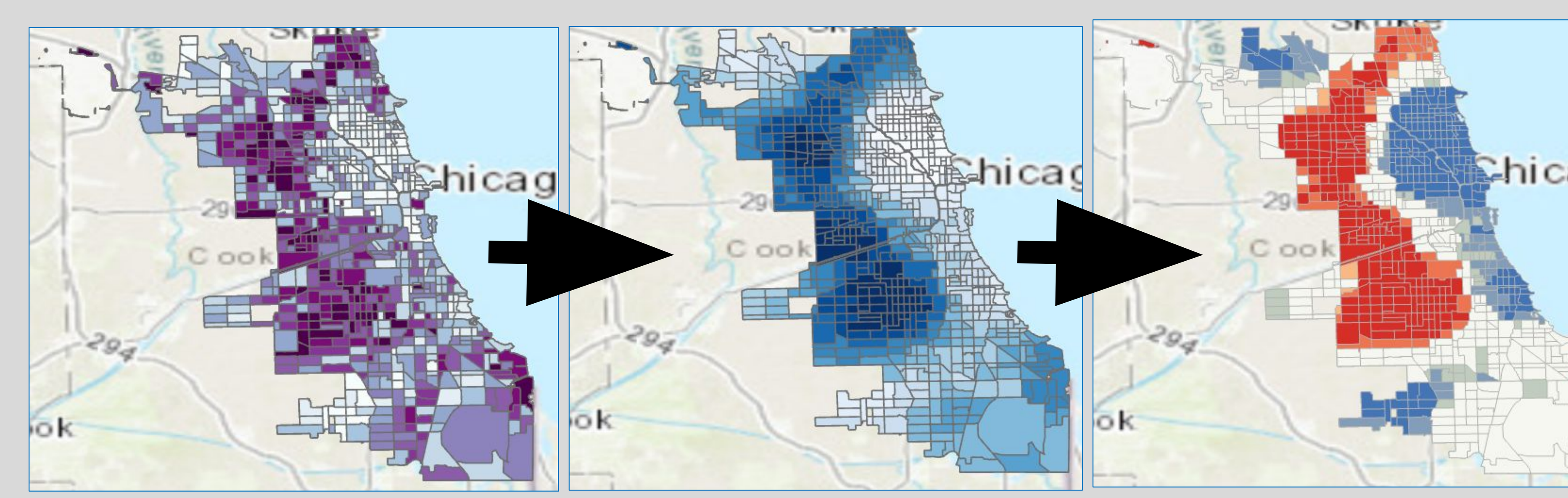
- Neighbors for each tract are prescribed using the fixed distance method. A set distance from the center of each feature is determined to construct assign neighbors such that no one feature has too many or too few. For spatial analysis, a minimum of 8 neighbors should be assigned to each feature, as recommended by Esri's suggested hotspot analysis practices [2]. A maximum number of neighbors is set on a case-by-case basis, so a maximum of 100 neighbors per tract was set. At the determined distance, ~1% of the tracts had less than 8 neighbors and all were under 100.

Prioritization Model

Our prioritization method is adapted from a model used by Safe Water Engineering in Washington, D.C. [7]. Replacement priorities are assessed at the block group level, which typically includes between 600 and 3,000 residents. Each block group is evaluated using two key metrics:

- Cost of Lead (CoL): A social harm score that reflects the potential impact of lead exposure on vulnerable populations. It incorporates the Area Deprivation Index (ADI), which is a composite measure of socioeconomic disadvantage, and the number of children under 5, who are most at risk from lead exposure
- Likelihood of Lead (LoL): The average number of service lines within the block group that are suspected to require replacement.

Each score is normalized on a scale of 1 to 10. By multiplying CoL and LoL, we generate a composite priority score ranging from 0 to 100, identifying which block groups should be prioritized for service line replacement.



Sample Hotspot Manipulation for % Uninsured Population: Raw Data, Z-scores, Confidence Intervals

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}}$$

Getis-Ord Gi* Statistic [1]

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \quad (3)$$

ACKNOWLEDGEMENTS

Assessing Chicago’s Lead Risk Against Varying Temperature

Gavin Coffer, Pranav Kuchibhotla, Thailer Simmons

Advisor: Robert Ellis (College of Computing, Illinois Institute of Technology), Subject Matter Expert: Elin Betanzo, P. E. (Safe Water Engineering)

METHODOLOGICAL APPROACH

This investigation begins by loading just the sample date, “Max,” and “1st Draw” lead measurements, converting dates to a proper datetime type, and then mapping each month to a season (Winter, Spring, Summer, Fall) in Python. Any readings at or above 1000 ppb are filtered out, keeping both summary statistics and plots focused on the realistic range of first‑draw concentrations.

Next, the script applies nonparametric bootstrapping to each season’s data: it resamples with replacement 10,000 times to build empirical distributions for both the mean and the 90th percentile and eliminate bias in the data. From those bootstrap distributions it extracts point estimates plus 95% confidence intervals (the 2.5th and 97.5th percentiles), giving robust, assumption‑free uncertainty bounds even when the underlying data are skewed.

Finally, it calculates each season’s mean and 90th percentile lead level and renders a log‑scale boxplot colored by season. The use of a logarithmic y-axis (ticks at 1, 10, 100, 1000 ppb) spreads out orders of magnitude evenly, giving a concise yet comprehensive view of how peak lead exposures vary throughout the year.

CONCLUSIONS

The box plot clearly shows a seasonal pattern in first‑draw lead levels:

- For the **max** lead draw: **Summer** samples average **6.97 ppb** (95 % CI 6.60–7.37), and **Fall** samples average **6.83 ppb** (95 % CI 6.36–7.34), both substantially higher than the colder-month means: **Winter** at **5.74 ppb** (95 % CI 5.27–6.26) and **Spring** at **5.79 ppb** (95 % CI 5.46–6.16).
- For the **first** lead draw: **Summer** samples average **4.58 ppb** (95 % CI 4.26–4.93), and **Fall** samples average **4.59 ppb** (95 % CI 4.17–5.07), both slightly higher than the colder‑month means: **Winter** at **4.32 ppb** (95 % CI 3.88–4.83) and **Spring** at **3.81 ppb** (95 % CI 3.53–4.14).
- On the log-scale y-axis you can see that, although outliers extend up toward 100 ppb (and a few up near 1000 ppb before filtering), the bulk of Summer and Fall samples lie above those of Winter/Spring, both in their box midlines (medians) and in the upper quartile.

When you look at the raw data, this makes sense: warmer water accelerates pipe corrosion and dissolved-lead release. June 1–September 30 is exactly the EPA’s required sampling window under the Lead & Copper Rule, because those are the “worst-case” months. Systems must collect all their first‑draw taps in that span and report the 90th-percentile lead result for compliance. By design, that period captures the true peak in Summer (and into early Fall) readings, so if only Winter or Spring were monitored, there’d risk missing an exceedance of lead levels.

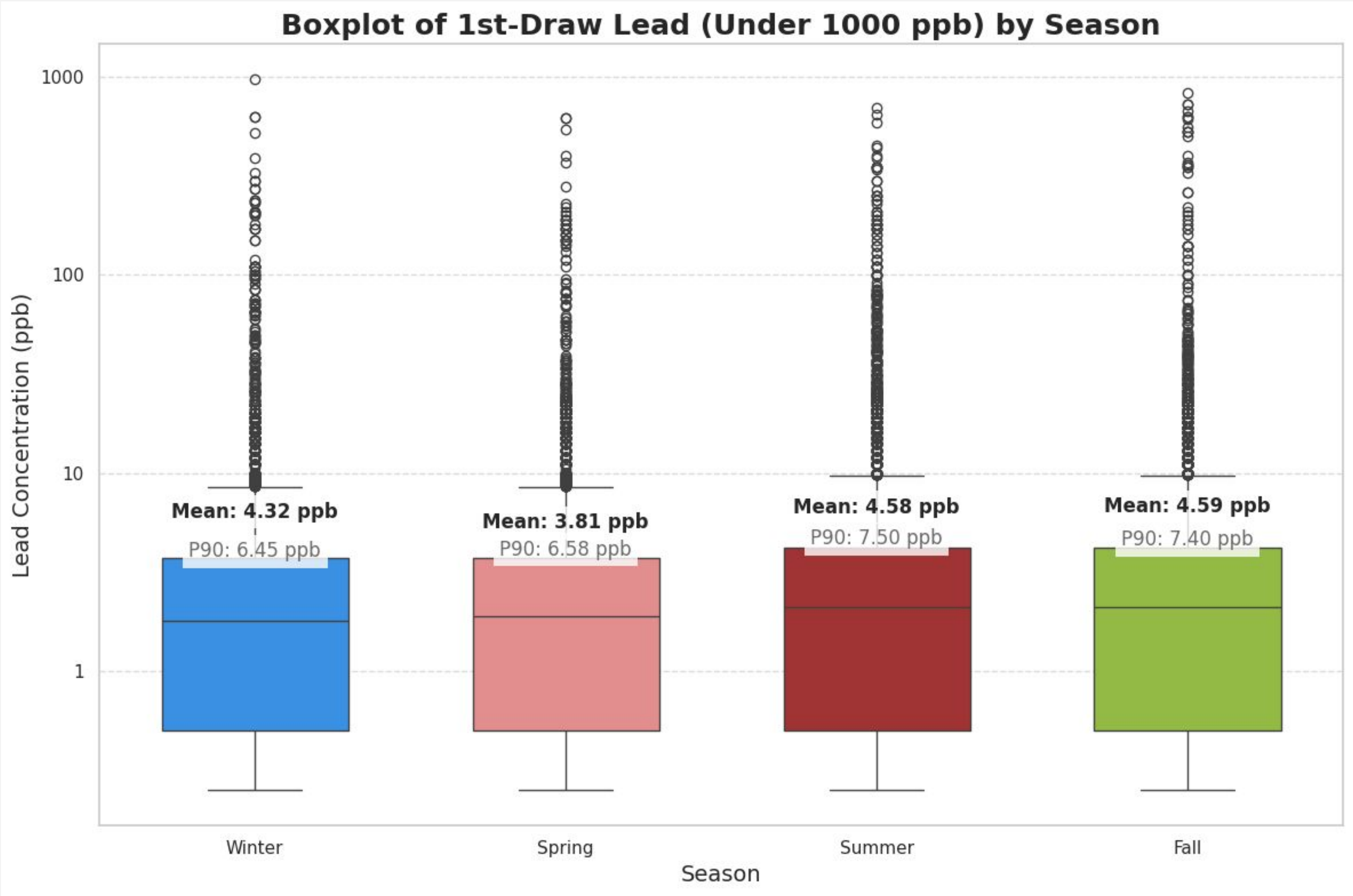
WHY DO LEAD LEVELS INCREASE IN HIGHER TEMPERATURE?

Higher temperatures accelerate the chemical reactions that corrode leaded materials. Reaction rates are found to roughly double for every 10 °C increase showing that warm water eats away at pipes and fixtures faster. At the same time, heat makes protective mineral scales (like lead carbonate or phosphate films) more soluble, causing both dissolved lead and particulate flakes to fall off into the flow. Disinfectant residuals (chlorine or chloramine) also decay more rapidly in hot water, shifting pH conditions toward greater metal release. Finally, longer stagnation times in summer (faucets unused for hours) give corrosion extra time to progress unchecked, all combining to push lead levels higher in the warm months.

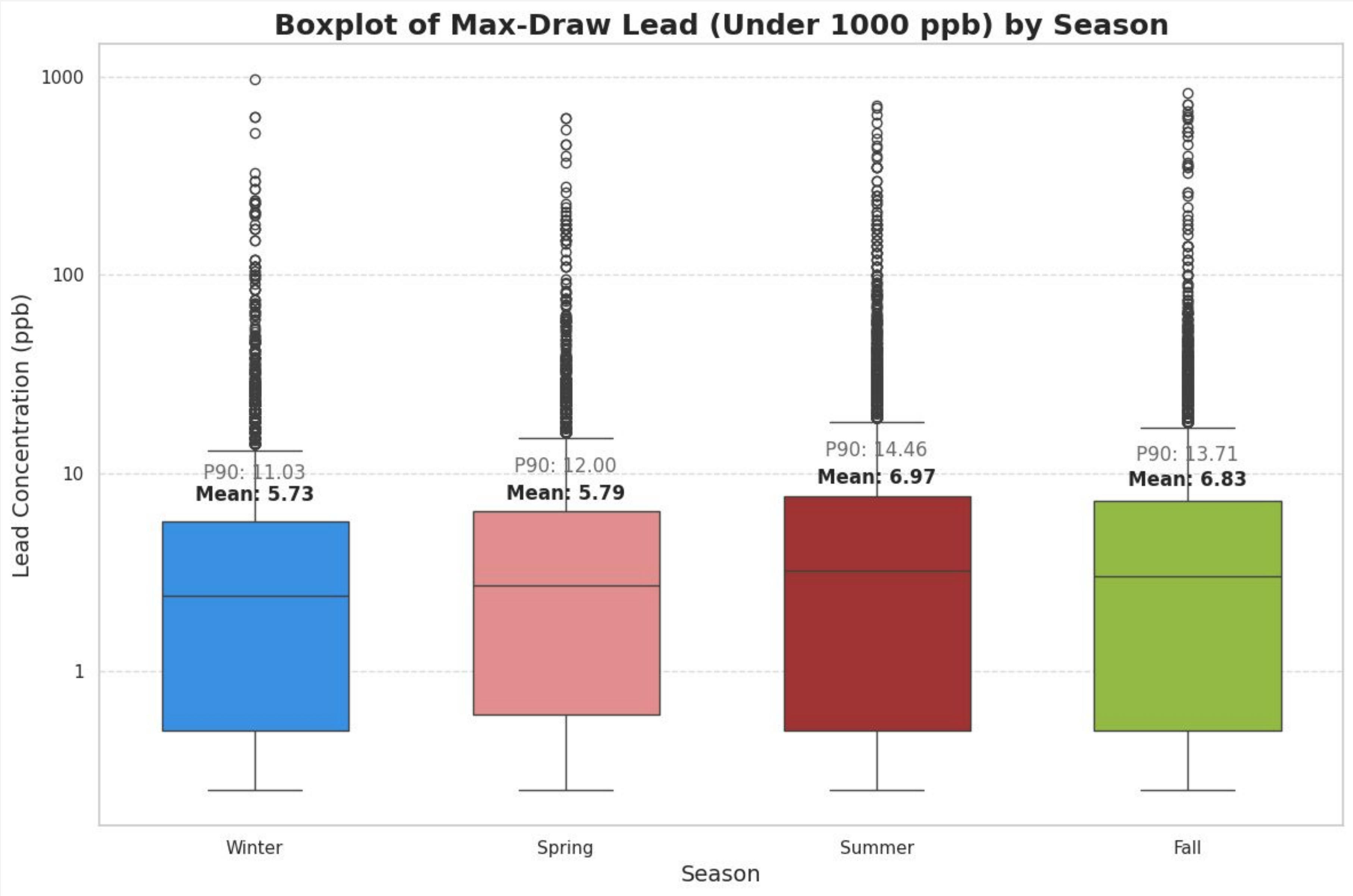
LEAD LEVEL DATABASE EXAMPLE

Sample Date	Address	1st Draw	2/3 Min	5 Min
1/27/2016	45XX N MERRIMA	3.22	2.17	1.94
4/2/2016	35XX W 84TH PL	2.67	2.11	3.56
7/6/2016	13XX E 52ND ST	4.5	9.7	7
9/21/2016	13XX E 55TH PL	4.7	1.6	1.4

1ST DRAW



MAX DRAW



MONTHLY TEMPERATURE DRIVEN LEAD TRENDS

Goal: Explore whether lead levels track changes in outdoor temperature across the years, beyond broad seasonal averages.

Methodology:

- Plotted 90th-percentile lead levels for each month (Max and First Draw) alongside average monthly temperature.
- Performed a permutation test comparing observed vs random alignment using Euclidean distance.
- 1,000 permutations were conducted to create a null distribution.

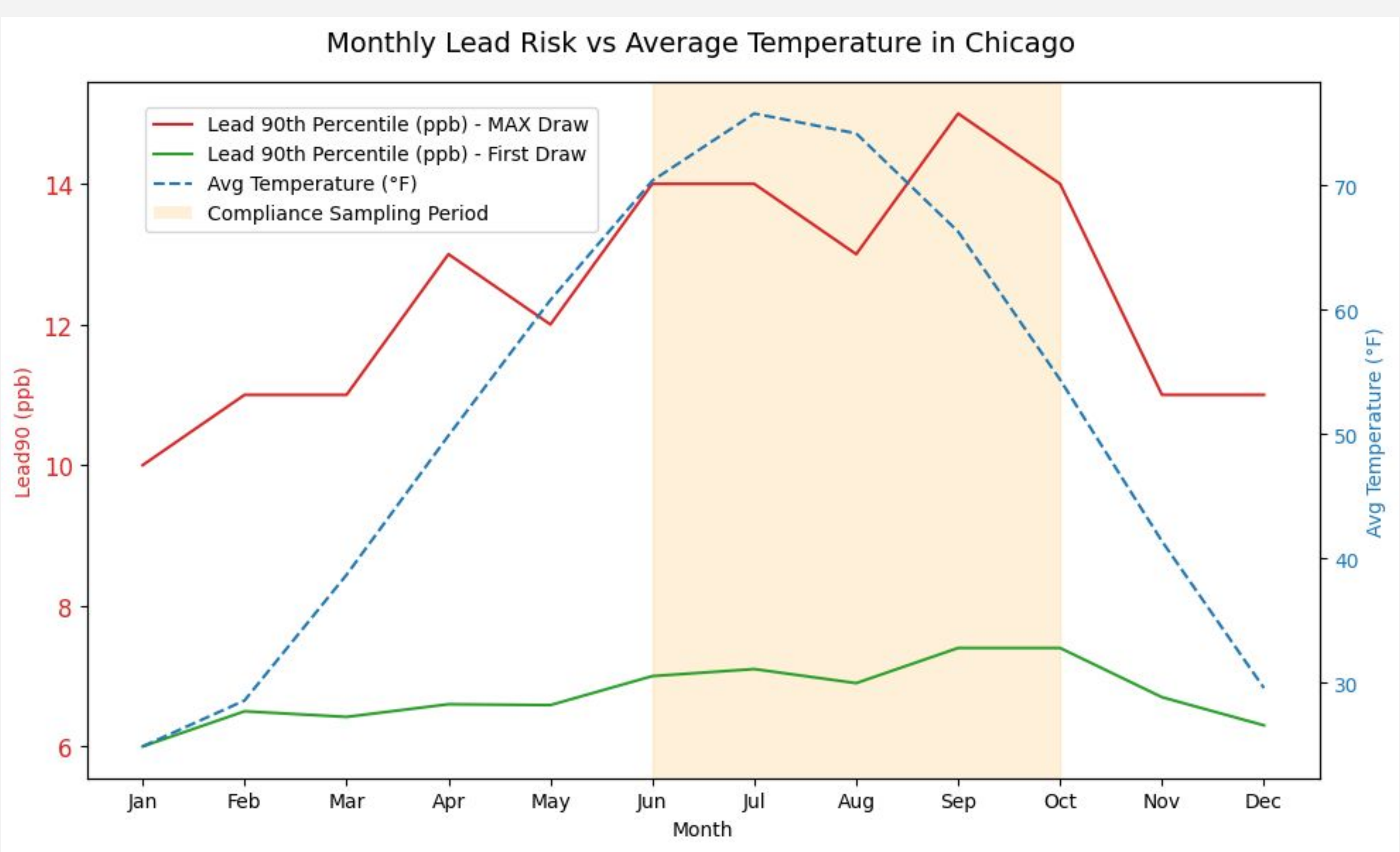
Key Results:

- Lead levels lag temperature by approximately 1 to 2 months, peaking in September–October after temperature peaks in July–August.
- 90th percentile of Max Draw lead levels peaks around 15 ppb; 90th percentile of First Draw peaks at 7 ppb.
- Observed Euclidean Distance = 146.02; p-value $< 6.10*e^{-4}$ → Strong statistical evidence of correlation.
- Lead risk rises after sustained warm periods, not immediately with temperature spikes.

Interpretation:

- Warmer temperatures may gradually increase corrosion and lead release into water supply systems.
- Seasonal public health alerts should focus on early to mid-summer, ahead of when lead levels peak in late summer.

MONTHLY LEAD RISK VS AVERAGE TEMPERATURE



REFERENCES

A. Sumner, T. (2019, August 8). *When measuring lead in water, check the temperature*. Science News. https://www.sciencenews.org/article/when-measuring-lead-water-check-temperature

B. Masters, S., Welter, G. J., & Edwards, M. (2016). Seasonal variations in lead release to Potable Water. *Environmental Science & Technology*, *50*(10), 5269–5277. https://doi.org/10.1021/acs.est.5b05060

C. Mohammadzadeh, M., Basu, O. D., & Herrera, J. E. (2015). Impact of water chemistry on lead carbonate dissolution in drinking water distribution systems. *Journal of Water Resource and Protection*, *07*(05), 389–397. https://doi.org/10.4236/jwarp.2015.75031

Assessing Chicago’s Lead Risk Against Varying Temperature

Gavin Coffer, Pranav Kuchibhotla, Thailer Simmons

Advisor: Robert Ellis (College of Computing, Illinois Institute of Technology), Subject Matter Expert: Elin Betanzo, P. E. (Safe Water Engineering)

METHODOLOGICAL APPROACH

This investigation begins by loading just the sample date, “Max,” and “1st Draw” lead measurements, converting dates to a proper datetime type, and then mapping each month to a season (Winter, Spring, Summer, Fall) in Python. Any readings at or above 1000 ppb are filtered out, keeping both summary statistics and plots focused on the realistic range of first‑draw concentrations.

Next, the script applies nonparametric bootstrapping to each season’s data: it resamples with replacement 10,000 times to build empirical distributions for both the mean and the 90th percentile and eliminate bias in the data. From those bootstrap distributions it extracts point estimates plus 95% confidence intervals (the 2.5th and 97.5th percentiles), giving robust, assumption‑free uncertainty bounds even when the underlying data are skewed.

Finally, it calculates each season’s mean and 90th percentile lead level and renders a log‑scale boxplot colored by season. The use of a logarithmic y-axis (ticks at 1, 10, 100, 1000 ppb) spreads out orders of magnitude evenly, giving a concise yet comprehensive view of how peak lead exposures vary throughout the year.

CONCLUSIONS

The box plot clearly shows a seasonal pattern in first‑draw lead levels:

- For the **max** lead draw: **Summer** samples average **6.97 ppb** (95 % CI 6.60–7.37), and **Fall** samples average **6.83 ppb** (95 % CI 6.36–7.34), both substantially higher than the colder-month means: **Winter** at **5.74 ppb** (95 % CI 5.27–6.26) and **Spring** at **5.79 ppb** (95 % CI 5.46–6.16).
- For the **first** lead draw: **Summer** samples average **4.58 ppb** (95 % CI 4.26–4.93), and **Fall** samples average **4.59 ppb** (95 % CI 4.17–5.07), both slightly higher than the colder‑month means: **Winter** at **4.32 ppb** (95 % CI 3.88–4.83) and **Spring** at **3.81 ppb** (95 % CI 3.53–4.14).
- On the log-scale y-axis you can see that, although outliers extend up toward 100 ppb (and a few up near 1000 ppb before filtering), the bulk of Summer and Fall samples lie above those of Winter/Spring, both in their box midlines (medians) and in the upper quartile.

When you look at the raw data, this makes sense: warmer water accelerates pipe corrosion and dissolved-lead release. June 1–September 30 is exactly the EPA’s required sampling window under the Lead & Copper Rule, because those are the “worst-case” months. Systems must collect all their first‑draw taps in that span and report the 90th-percentile lead result for compliance. By design, that period captures the true peak in Summer (and into early Fall) readings, so if only Winter or Spring were monitored, there’d risk missing an exceedance of lead levels.

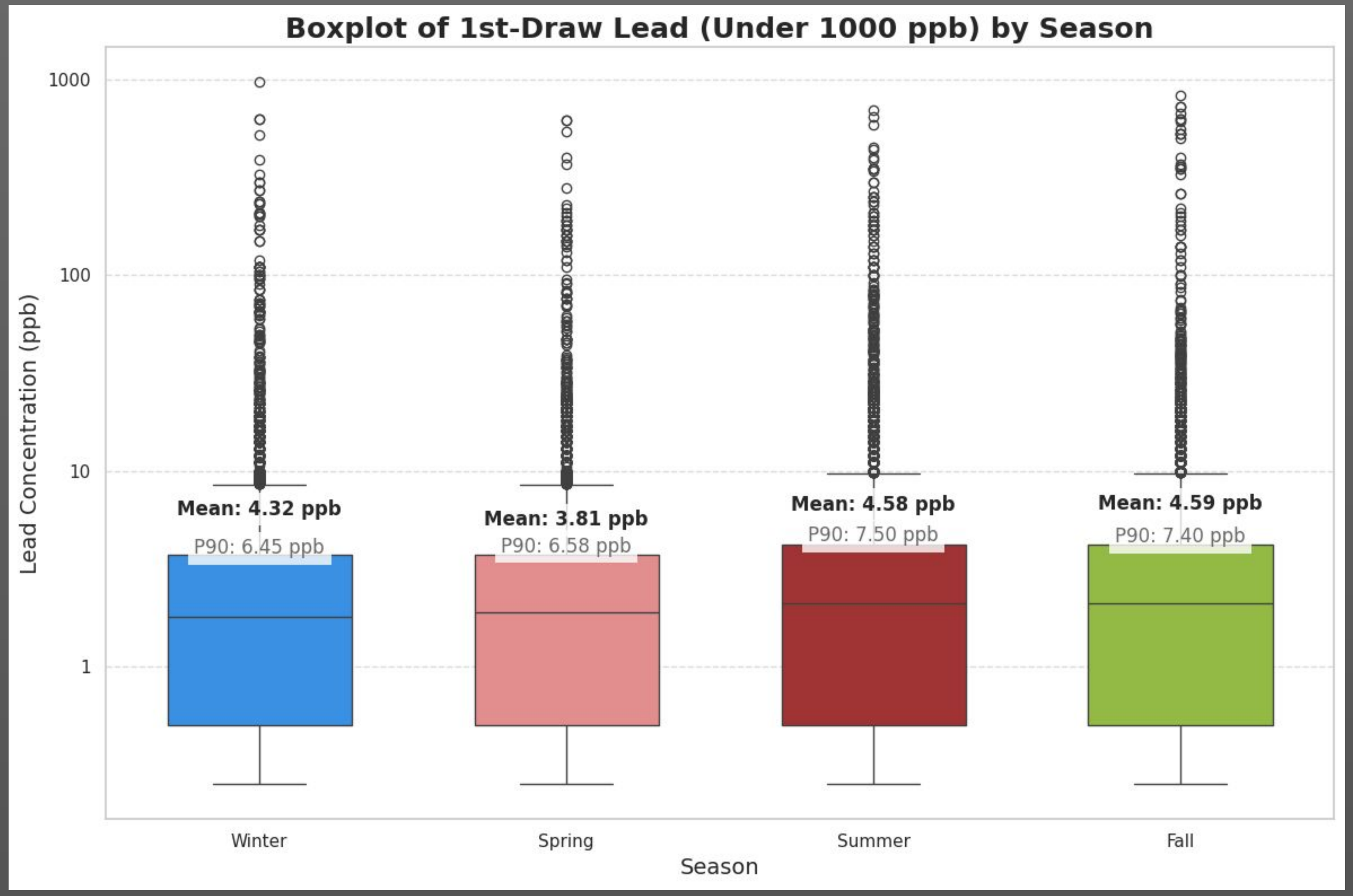
WHY DO LEAD LEVELS INCREASE IN HIGHER TEMPERATURE?

Higher temperatures accelerate the chemical reactions that corrode leaded materials. Reaction rates are found to roughly double for every 10 °C increase showing that warm water eats away at pipes and fixtures faster. At the same time, heat makes protective mineral scales (like lead carbonate or phosphate films) more soluble, causing both dissolved lead and particulate flakes to fall off into the flow. Disinfectant residuals (chlorine or chloramine) also decay more rapidly in hot water, shifting pH conditions toward greater metal release. Finally, longer stagnation times in summer (faucets unused for hours) give corrosion extra time to progress unchecked, all combining to push lead levels higher in the warm months.

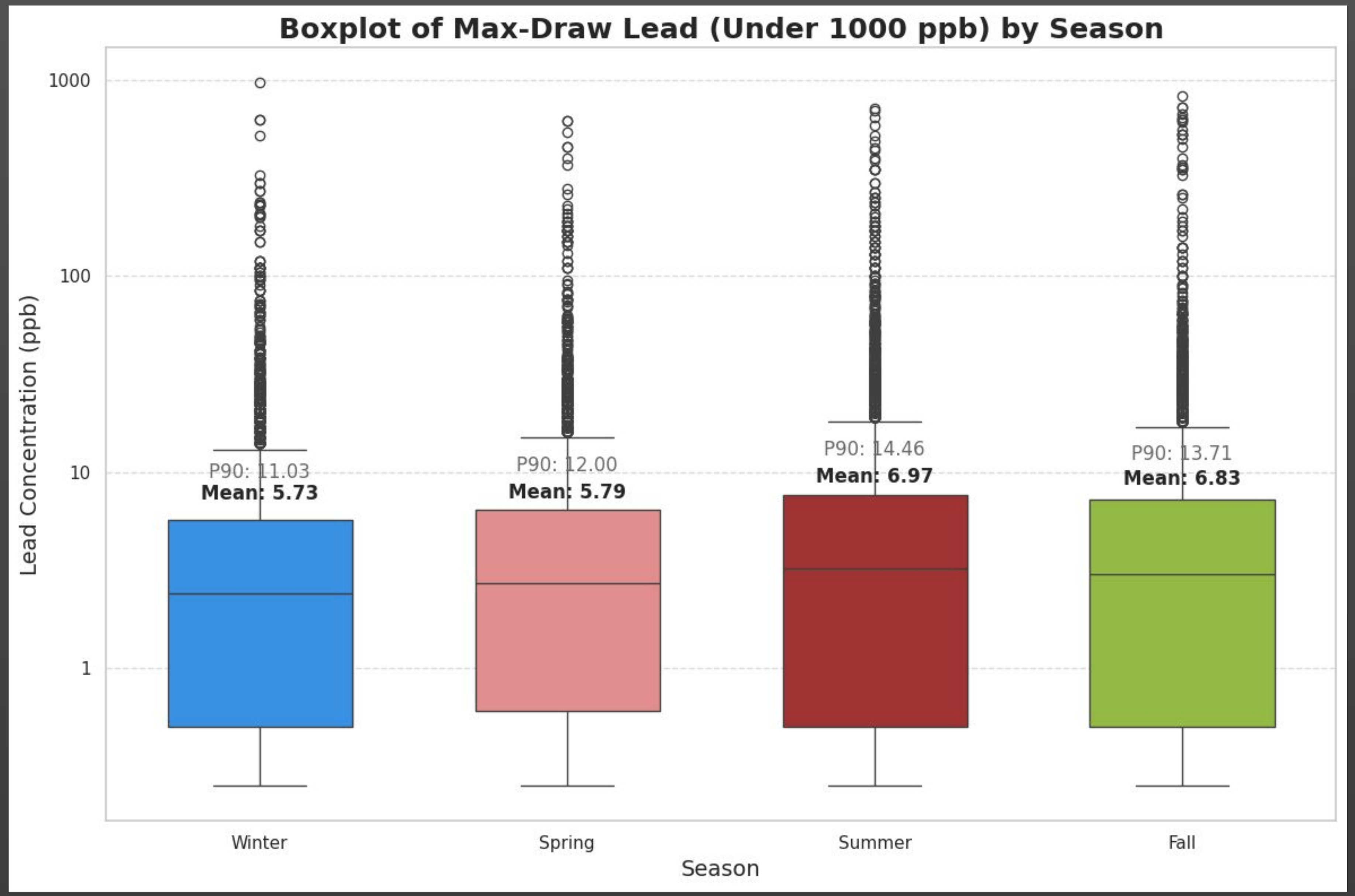
LEAD LEVEL DATABASE EXAMPLE

Sample Date	Address	1st Draw	2/3 Min	5 Min
1/27/2016	45XX N MERRIMA	3.22	2.17	1.94
4/2/2016	35XX W 84TH PL	2.67	2.11	3.56
7/6/2016	13XX E 52ND ST	4.5	9.7	7
9/21/2016	13XX E 55TH PL	4.7	1.6	1.4

1ST DRAW



MAX DRAW



MONTHLY TEMPERATURE DRIVEN LEAD TRENDS

Goal: Explore whether 90th-percentile lead levels in Chicago water track seasonal patterns in outdoor temperature.

Methodology:

Plotted monthly 90th-percentile lead levels(Lead90) (Max and First Draw) alongside average monthly temperature. Tested the alignment using a permutation-based Euclidean distance test:

- **Null hypothesis:** No meaningful relationship between average monthly temperature and 90th-percentile lead levels, any similarity is due to random chance.
- **Alternative**: Temperature and lead levels follow a similar pattern more closely than expected by chance.
- **Test statistic:** Euclidean distance between normalized Lead90 and temperature curves (measures shape similarity).
- Shuffled Lead90 values 1,000 times to build a null distribution.

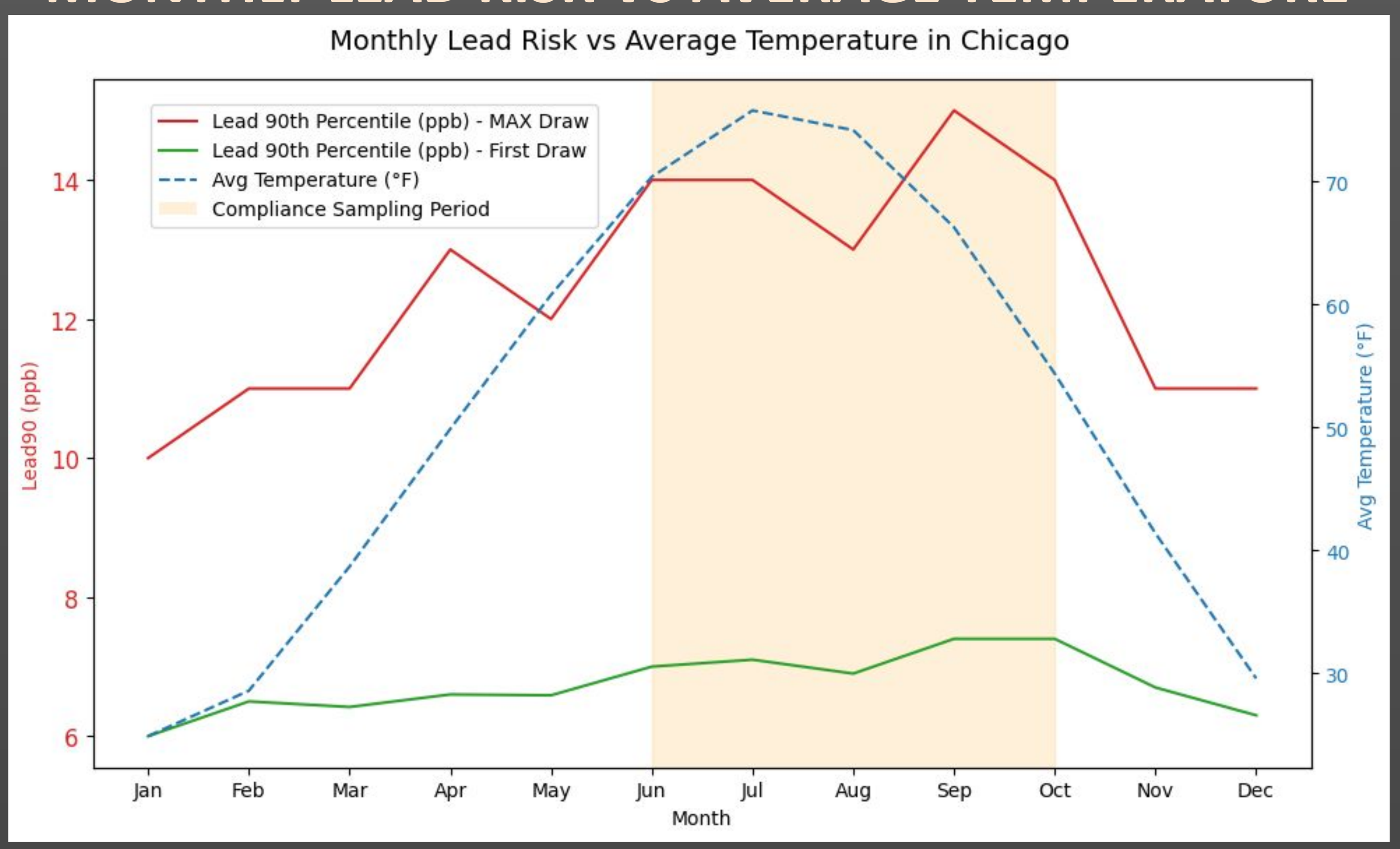
Key Results:

- Observed distance = 146.02
- p-value < 6.1e-4 → Reject null hypothesis
- Max Draw 90th-percentile peaks ~15 ppb; First Draw ~7 ppb (in Sept–Oct)
- Lead levels peak 1–2 months after temperature peaks (July–August → Sept–Oct)

Interpretation:

- Sustained warm weather may increase corrosion and lead release.
- Public health alerts should begin in early summer, before late-summer peaks.

MONTHLY LEAD RISK VS AVERAGE TEMPERATURE



REFERENCES

A. Sumner, T. (2019, August 8). *When measuring lead in water, check the temperature*. Science News. https://www.sciencenews.org/article/when-measuring-lead-water-check-temperature

B. Masters, S., Welter, G. J., & Edwards, M. (2016). Seasonal variations in lead release to Potable Water. *Environmental Science & Technology*, *50*(10), 5269–5277. https://doi.org/10.1021/acs.est.5b05060

C. Mohammadzadeh, M., Basu, O. D., & Herrera, J. E. (2015). Impact of water chemistry on lead carbonate dissolution in drinking water distribution systems. *Journal of Water Resource and Protection*, *07*(05), 389–397. https://doi.org/10.4236/jwarp.2015.75031

Chicago has more lead service lines than any other U.S. city (over 400,000) posing a major public health concern. This project prioritizes replacement efforts using a block group–level risk model that combines demographic vulnerability (social cost of lead) with infrastructure data (likelihood of lead). Hotspot analysis shows that West and South Side neighborhoods, particularly Englewood and Austin, face the greatest risk due to both a high concentration of lead lines and higher vulnerability among residents. In contrast, Central and Northern areas show lower risk. Our model helps identify where lead service line replacement will have the greatest health and equity impact.

Children under 5 ACS data:

https://data.census.gov/table/ACSDT5Y2022.B01001?q=B01001&g=050XX00US17031$1500000,17043$150000

ADI rankings

https://www hoodatlas.medicine.wisc.

edu/downlo

CoL Score
CoL = 0.5 * ADI Score + 0.5 * Children under 5 score

Children under 5 score
50+ children: 10
11-50 children: 5
1-10 children: 3
No children: 1

LoL score
Lead and galvanized lines: 10.
• Unknowns: 5

Averaged for block group

Model score = CoL * LoL

CoL Score
CoL = 0.5 * ADI Score + 0.5 * Children under 5 score

Children under 5 score
50+ children: 10
11-50 children: 5
1-10 children: 3
No children: 1

Lol Score
Lead and galvanized lines: 10.
Unknowns: 5

Averaged for block group

Model score = CoL * LoL