DS 100/Math 100 Project: Locating Hotspots for Lead Contamination of Chicago Drinking Water

Introduction

Lead contamination in drinking water is a serious public health concern, particularly in cities like Chicago where lead was ubiquitously used until it was banned in 1986. Elevated lead levels have been detected in thousands of tests of household drinking water across the city, with potentially around 412,000 water service lines containing lead (see [1]). See also a recent article from *The Guardian* [4]. The US Environmental Protection Agency regulates the amount of lead in drinking water through the Lead and Copper Rule (LCR, [2]) and the Lead and Copper Rule Improved (LCRI, [3]), which started taking effect in October, 2024. A key component of the LCR is that cities like Chicago must test household water, and enforcement actions follow if 10% of the households tested meet or exceed 15ppb of lead. The LCRI will reduce the trigger threshold from 15ppb to 10ppb.

The LCR and even the LCRI are significantly deficiency in protecting Chicago residents from lead in many ways, including:

- Children are adversely affected by much lower levels of lead than 10ppb;
- 9.9% of households could have any level of lead without triggering enforcement or even notification of residents; and
- High lead exposure can be concentrated in a city block or small neighborhood without triggering enforcement or notification.

Project main goals

The suggested goal of this project is to find a neighborhood 'hotspot' of elevated lead levels in Chicago—such as a census tract, contiguous grouping of census tracts, or a neighborhood—as determined by analysis of test results published by the Chicago Department of Water Management. "Suggested" means you may propose a modification to the project that is mutually agreeable to the team and the instructor. The main data sources for estimating lead risk are lead test data at Chicago Department of Water Management, and the Chicago service line inventory (see also the more user-friendly Sun-Times/WBEZ version). Methodology can be drawn from descriptive statistics, statistical tests, and geographic information techniques for data visualization, kernel-density methods, or other approaches, to help understand patterns in the data and their implications for public health. Specific subgoals to complete are:

- **Read!** See the WBEZ webpage for the reporter Keerti Gopal from Inside Climate news here, which links to multiple recent articles on lead service lines in the City of Chicago. Her main WBEZ coauthor on this project is Juanpablo Ramirez-Franco.
- **Data review.** Download the dataset from the Department of Water Management here. Review the contents to understand the structure of 3-draw and sequential data and the meaning of each field (e.g., address, lead measurements). Read the website itself (https://chicagowaterquality.org/) to understand how data was collected. (Most likely you will need to use a Python package such as `datascience` or `pandas`.) Explore and understand the

Chicago service line inventory data set (or Sun-Times/WBEZ version), including in the context of map visualization.

- **Descriptive statistics**. Summarize the distribution of lead levels using mean, median, standard deviation, histograms, and boxplots. Compare sequential draw data with 3-draw data to see differences in lead levels. Focus particularly on the 1st draw of the 3-draw data, which is most relevant for compliance with the LCR and LCRI.
- **Hotspot identification.** Based on the street addresses of drinking water tests, identify a small local region of the city of at least 1000 residents where the tested lead levels are much higher than the city as a whole. A Chicago census tract is a good target size; a Chicago neighborhood is a more challenging size but also interesting.
- **Bootstrap hypothesis testing.** Conduct an empirical hypothesis test using bootstrap sampling to determine statistical significance of the elevated levels of lead in your selected region compared to the rest of the city.
- **Statistical testing.** Conduct a statistical test such as a t-test to determine statistical significance of the elevated levels of lead in comparison to the rest of the city.
- **Demographic inspection.** Look up demographic information on the residents of the region you identify—this is straightforward for a census tract, for example. Summarize demographic characteristics of the region's population, such as ethnicity, socioeconomic status, and especially number of children in each age range.
- **Summary and reporting.** Summarize and report on both your analytical findings and the vulnerability of the residents of the region to lead exposure, including regarding the number of children affected, and how lead levels compare to the guidelines in the LCR and LCRI.

Note: You will have flexibility on various aspects of how you approach the project, in consultation with your project mentor, Robert Ellis (ellisr@illinoistech.edu).

You are also encouraged to pursue the following "stretch" goals, or personalize your project with your own goals!

Optional "stretch" goals

1. **Exploratory data analysis.** Examine the variation in typical lead levels results across several census tracts, neighborhoods, or other type of region, focusing on the various "hotspots."
2. **Geocoding.** Convert the addresses of the 3-draw data into geographic coordinates (latitude and longitude) using geocoding tools. Plot the testing sites on a map of Chicago, or perhaps just in the local region you select.
3. **Hotspot plotting.** Create a heat map, such from putting kernel functions at geocoded address locations where high lead levels were found, or from counts of high test results within a grid, or similar.
4. **Expanded impact estimation.** Look at more or even all regions across Chicago to summarize the risk of lead in drinking water, especially to children, but potentially also to demographic groups that have more barriers to mitigating exposure to lead on their own, such as due to economic or language issues.

References

[1] Ramirez-Franco, J., Gopal, K., Aldhous, P., Aldern, C., & Qin, A. (2025, August 28). *How we mapped Chicago's lead service lines and what we learned*. *Chicago Sun-Times.* Retrieved September 25, 2025, from https://chicago.suntimes.com/environment/2025/08/28/map-chicago-lead-service-line-water-equity-what-we-learned

[2] U.S. Environmental Protection Agency. (n.d.). *Lead and Copper Rule* (EPA). Retrieved September 25, 2025, from https://www.epa.gov/dwreginfo/lead-and-copper-rule

[3] U.S. Environmental Protection Agency. (n.d.). *Lead and Copper Rule Improvements (LCRI)* (EPA). Retrieved September 25, 2025, from https://www.epa.gov/ground-water-and-drinking-water/lead-and-copper-rule-improvements

[4] McCormick, E. (2024, March 18). *Nearly 130,000 children exposed to lead-tainted drinking water in Chicago*. The Guardian. Retrieved September 25, 2025, from https://www.theguardian.com/us-news/2024/mar/18/chicago-children-lead-drinking-water-study