

Lead Exposure Project: Checkpoint 1

Katherine Laliotis, Alex Schimmoller, Brock Grafstrom

May 23, 2025

1 Datasets

For this project, we will primarily use two datasets.

First, the [Water System Service Line Inventory](#) from the City of Columbus contains information about the type of water line used in homes. For the homes and buildings included in this dataset, the data includes the type of material used in the line, and a lead-Non-lead classification.

Second, we will use data from the [Franklin County Auditor's Office](#) on home ages in the area. The auditor's office data base contains addresses, homeowner names, tax and sale information, year built, area, and rooms and bathrooms. We propose to start by using year built for houses, but we are interested in using locations or school district designations to compare homes in different neighborhoods as well.

2 Problem

Lead can damage nearly every system in the human body, and has harmful effects on both adults and children. It is a serious environmental public health threat to people across the United States, including those living in Central Ohio. Oftentimes, lead exposure takes place within one's own home through ingestion of water carried through lead piping. In an effort to raise awareness of citizens' possible exposure to lead, the City of Columbus makes publicly available a Service Line Material Inventory detailing the type of line used for each home. However, this data is oftentimes self reported, and there exist many homes not registered with the inventory. In this project, we will use the inventory and data on homes available through the Franklin County Auditor's Website to train a model to determine a home's probability of containing lead piping. This model could then be used to assign unregistered homes with a relative risk of containing lead pipe.

3 Project Stakeholders

The following list includes potential stakeholders who might benefit from the results of this project:

- Columbus, Ohio residents, in particular those whose homes are not reported in the [Water System Service Line Inventory](#)
- The City of Columbus, whose current initiatives include the [Lead Education and Poisoning Prevention Program](#) and the [Lead Service Line Replacement Program](#).

4 Key Performance Indicators

Key performance indicators (KPIs) we will use for this project include:

1. Accuracy of the model in predicting homes that we know do contain lead using cross-validation and forward selection
2. Assuming the model proves to be accurate, results of this project could be used by the City of Columbus to create a list of all homes with the highest risk of lead exposure from water service lines and motivate the highest priority projects within the [Lead Service Line Replacement](#)

Program. As there is no known “safe” blood lead level above zero, successful implementation of this program can lead to the minimization of lead poisoning cases in Central Ohio.

3. Knowledge of the number of homes whose pipes need to be replaced would also help the City of Columbus to design an accurate and cost-effective plan to meet the EPA’s deadline of 10 years to replace lead pipes. Developing a plan and a cost estimate takes time, and throughout that time inflation and increasing materials costs would cause the replacement to only get more expensive for the city as time goes on. Using our findings, the city could minimize the impact of inflation on the pipe replacement project.