> ***We set out to research and optimize the allocation of a $4.2 million grant to reduce lead exposure within the City of Grand Rapids' homes.***

Overview

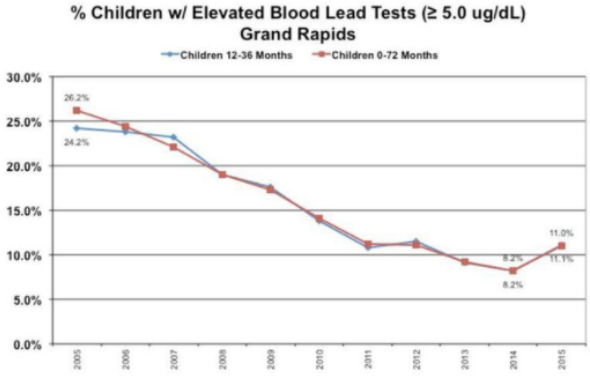
We partnered with the City of Grand Rapids to study the risk of lead-based paint in houses across Grand Rapids. The main questions we explored were:

1. What are the health risks of lead-based paint?
2. How much will removal of lead-based paint and outreach cost?
3. What have previous programs across the country done?
4. What can the current *Get The Lead Out!* program in Grand Rapids improve upon for lead-based paint remediation?

We found that in Grand Rapids, of the 55,000 known addresses, nearly 49,000 homes were built before 1978, the year lead-based paint was banned. Removal of lead from all of the houses built before 1978 would cost an estimated $735,002,406, which is far out of reach with the grant funding currently available. We used clustering algorithms to determine groups of houses that are of highest risk based on year built, assessed value, and other available metrics. Our goal is to reduce the total number of houses observed to be those of high risk with the funding available.

Background

Before 1978, lead was a primary material used in home construction, or more specifically in water pipes and paint. Lead-based paint was banned in the United States in 1978, but by then an estimated 37 million houses were built containing the harmful substance. As mentioned, Grand Rapids has about 49,000 homes built before 1978. This means approximately 90% of the homes in Grand Rapids are in danger of lead-based paint hazards. Grand Rapids is known for having a much higher rate of lead poisoning in children compared to the rest of Michigan.



The City of Grand Rapids received a $4.2 million grant to address the issue of lead-based paint in Grand Rapids. To address the issue, we created a mathematical model to best distribute that money. To create the model, we needed to determine the houses within the city that have the highest risk of lead-paint exposure, as well as create a cost model to divide the money reasonably throughout the city. We were asked to factor the city's Neighborhoods of Focus into our model. These are the neighborhoods within Grand Rapids that require extra attention and support.

Health Risks of Lead

Lead is harmful to people of all ages, but especially to children and infants. At lower levels of exposure, lead can affect brain development in children and cause behavioral changes such as lowered attention span and IQ, as well as create tendencies toward antisocial behavior. At high levels of exposure, lead can cause damage to the brain and the central nervous system resulting in convulsions, comas, and possibly death. Lead is also believed to cause immunotoxicity and damage to renal and reproductive systems. At any level of exposure, the neurological and behavioral effects are irreversible. Infants and unborn children can be exposed to lead through their mother. Any lead absorbed by the mother can be passed to the child through the placenta or through breastfeeding. Unborn children exposed to lead may result in premature birth, miscarriage, spontaneous birth, an

underweight birth, and neurological effects. It is also important to note that children and infants absorb more ingested lead than adults and they tend to show more symptoms at lower levels of exposure than adults.

Testing for Lead Paint in Homes

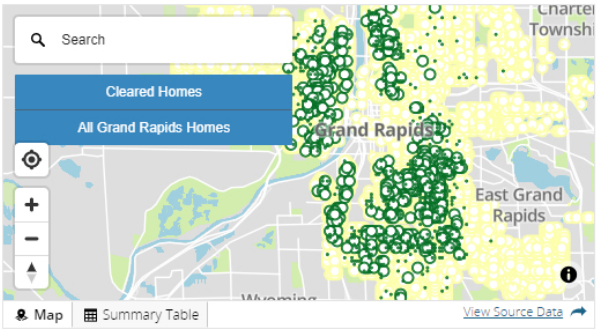
Once lead-based paint is identified within a home, a decision needs to be made to determine which method is the best way to deal with it. One of the cheaper methods of handling lead-based paint is encapsulation. This involves using a special sealing paint to trap the lead-based paint underneath and costs about $50 per gallon or about $800-$1400 per house. Something to keep in mind with this strategy is that it must be periodically inspected to ensure the seal is still air-tight. Another technique for interim control is enclosure which is very similar to encapsulation. However, instead of using a paint to seal away the lead-based paint, a dust-tight material such as drywall or aluminum is used to cover the old surface. This technique can vary greatly in price depending on the covering material used, but generally costs $10 per square foot. During this construction, workers must be very careful not to disturb the lead-based paint and have dust circulating in the home. Lastly, and of course the most expensive, is lead-paint abatement. This removal process can cost up to $20,000 depending on the size of the house and very careful measures must be taken to prevent lead paint dust from circulating. Each homeowner must determine the best and most affordable method for their house.

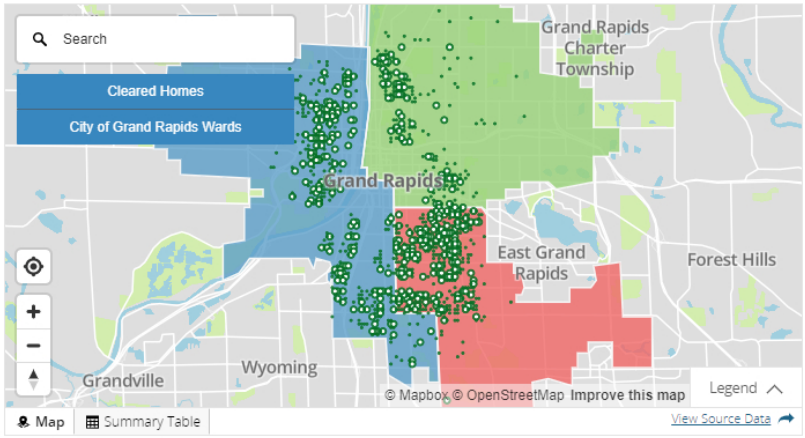
Analyzing the Lead Risk in Grand Rapids

In order to approach this problem, we collected the following datasets:

- **Registry Data** - Collection of all housing addresses in Grand Rapids with year built, assessed value, and home address. There are over 55,000 houses on this list.
- **Census Data** - Includes demographic metrics for each census tract in Grand Rapids from the City Health Dashboard.
- **Cleared Homes** - A list of 1600 homes that the Grand Rapids' *Get The Lead Out!* program has already done some level of interim controls on.

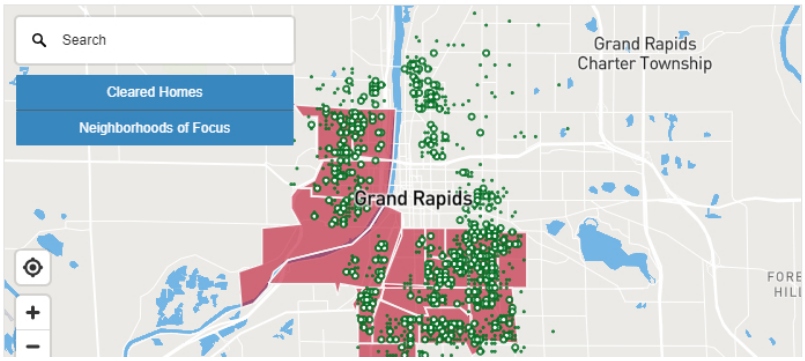
This map shows the cleared homes (green) as a layer on top of the Registry Data (yellow).

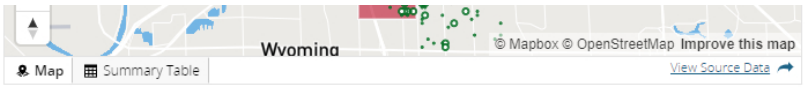


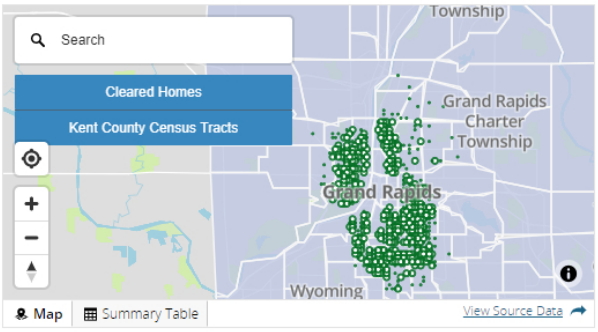


Focusing more on the locations of the cleared homes that have been helped in Grand Rapids, we put a layer with the City of Grand Rapids' wards underneath them. One goal with the $4.2 million grant is to distribute funds equitably across the three wards.

We next put a layer of the Neighborhoods of Focus underneath the cleared homes to see how much help they have received in comparison to the rest of the city. This showed us that they have received the majority of help from the city thus far which was great to see.





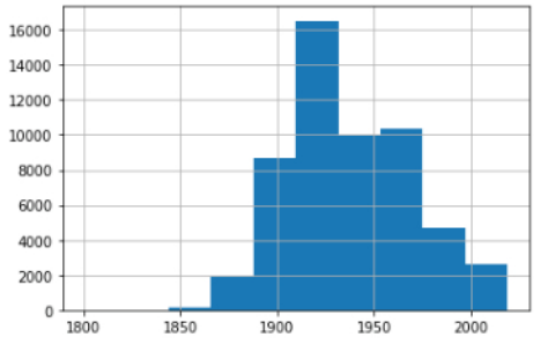


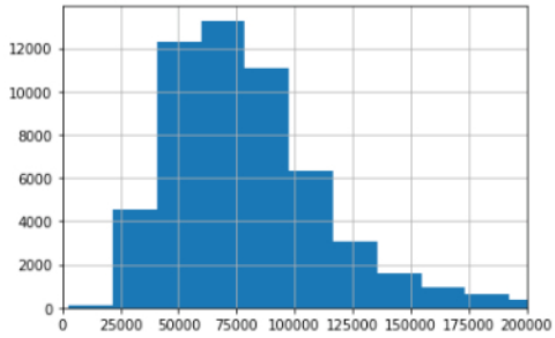
Lastly, we decided to map the cleared homes dataset on top of the Kent County Census Tract map. This is a great visual to see how much help each census tract has received thus far.

Clustering Model

The available data provided relevant information, such as the age, assessed value, and lead risk, for each home in Grand Rapids. To identify homes that are most likely to be in need of abatement, we determined which homes are most similar to the cleared homes that have previously been abated through the Get the Lead Out! program. Our strategy involved breaking the homes in Grand Rapids into about 20 clusters, where the homes within each cluster are all similar to one another. Because the majority of the cleared homes fell within a small number of clusters, we concluded that the other homes in those clusters are likely candidates for abatement.

To begin the clustering, our first clustering used the variables "Year Built", "2020 AV" (assessed value) and "Home Lead Risk". We began by taking the mean and standard deviation of these variables to get a sense of what we could consider good/bad for the resulting clusters. As we can see in the histograms below, a vast majority of the homes in Grand Rapids have been built before 1978 and are valued at below $100,000.



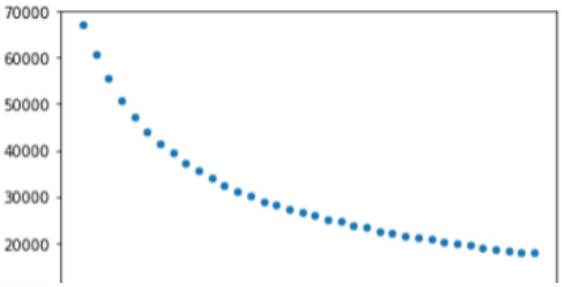


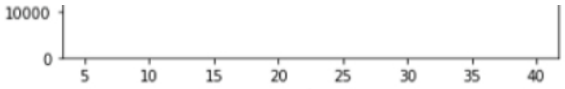
The computations show that the average year built is 1937, the average assessed value is $81,893 and the average home lead risk is 15%.

Next, we performed a standard normalization of the data so that each of the three variables was equally weighted when quantifying the similarity of homes. For example, if we just look at the numerical values themselves, the maximum assessed value of a house in 2020 is going to be over $400,000 and the maximum home lead risk is below 40%. Normalizing the data would ensure that these values would be weighted equally in our algorithm.

We decided to use the k-means clustering algorithm which is a fundamental machine learning algorithm implemented in the *scikit learn* Python package. A k-means clustering takes the three variables as inputs and groups houses that have similar values across the three variables. For example, if we ran a clustering using five clusters, we would end up with five groupings of houses. The question then becomes how many clusters should we use? We used a loop to run a k-means clustering with different numbers of clusters, varying from 5 clusters to 51 clusters.

Then we plotted the average distance of data points from the center of their respective cluster to determine a range of clusters that would be appropriate to run an effective clustering. During this process, we were looking for an elbow or a point on the plot where the graph seems to level out. At this elbow, it shows that adding more clusters becomes unnecessary. As you can see in the figure to the right, the plot begins to level off around 20 clusters.

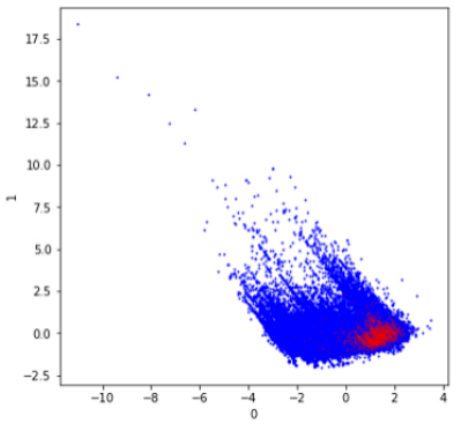




Analyzing the Clustering

After running the algorithm with the previously mentioned variables and 21 clusters, we get a clustering with several useful properties. First off, we can tell this is a good clustering because the clusters all represent a significant number of census tracts. This shows that our clusters are not geographically biased and allows us to focus solely on the properties of homes rather than where they are located. Our top three clusters-- clusters 10, 12, and 4 -- have above 4% of homes cleared, with the average year built before 1920, assessed values all below $50,000 and well above average home lead risk which is shown in the figure below. We can see from the clustering that as the percentage of homes abated per cluster goes down, the assessed value generally goes up along with the year the homes were built while home lead risk generally trends downward. This provides further justification for the use of this algorithm to detect homes most likely in need of abatement.

cluster	abated	total	% abated	Year Built	2020 AV	Home Lead Risk	# of census tracts	# of remodels	% remodeled
10	167	2724	0.061307	1896.856461	44631.057269	43.117988	22	620	0.227606
12	323	5480	0.058942	1919.252555	44669.288321	51.447847	27	852	0.155474
4	237	4286	0.055296	1893.696454	48317.032198	55.037704	21	949	0.221419

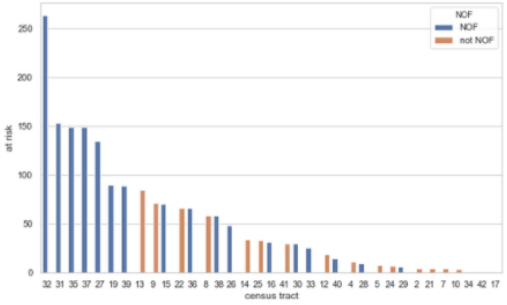


The graph of our principal component analysis for this clustering is seen in the figure on the left. The red points represent cleared homes while the blue points consist of the rest of the homes. We can see a pretty obvious clustering in the lower right corner of our graph which once again shows us that there is an obvious correlation between our variables and cleared homes. The information below tells us that as we move to the right along the x-axis of our graph, "Year Built" is decreasing, "2020 AV" is decreasing (at a slower rate, but still decreasing) and our "Home Lead Risk" is increasing at almost the same rate "Year Built" is decreasing! So the data points in the lower right hand corner of our graph are the homes that were built the earliest, are worth the least amount of money, and have been evaluated as the homes with the highest lead risk-therefore houses with similar characteristics should be considered for lead testing.

K-Nearest Neighbors

After we formed our clusters, we decided to run our top six clusters through another machine learning algorithm called the k-nearest neighbors algorithm. This algorithm finds the nearest data points to each point. It then outputs the percentage of the nearest data points that have had interim controls. For our algorithm, we used the 20 nearest neighbors. So if we have a home who has 3 of its nearest 20 data points with interim controls, this algorithm will output 15%. Since less than 5% of the total homes in the combined top 6 clusters, 15% is a significant number. We then decided to create two datasets to hand over to the city. One is our primary recommendations which consist of 1,825 addresses of homes who had 3 or more of their nearest 20 data points known to have had lead paint.

Again these are the homes we think should be the primary focus for the city as they are the most likely to contain lead-based paint on our analysis. The other dataset is a set of 3,703 homes who had exactly 2 of their nearest 20 data points known to have had lead paint. These homes should be the main focus of the city if there is money left after the previously mentioned 1,825 homes. The image to the right is a bar chart of the number of homes from the 1,825 homes from the k-nearest neighbors algorithm sorted by census tract. As you can see the blue bars represent NOF districts and the orange bars represent districts that are not considered NOF. It is worth noting that 75.89% of the 1,825 homes are from NOF districts.



Recommendations for the City of Grand Rapids

Wrapping up this semester of researching and modeling we have compiled a few recommendations for the City of Grand Rapids and more specifically, the *Get the Lead Out!* program. To start, the *Get the Lead Out!* program should begin collecting, keeping, and organizing all information/data about the homes they provide services for. This should include costs of various testing methods, interim controls, abatement, and general labor. With some additional work, as more data is compiled from the services that *Get the Lead Out!* performs, it could be used to adapt the cost model we created to make it more accurate and better service the city. Additionally, we recommend that the *Get the Lead Out!* program takes advantage of the detailed lead-based paint risk model we created. This model identifies specific addresses within Grand Rapids that have high-risk for lead-based paint and most likely need funding assistance. With this valuable information, the *Get the Lead Out!* program should use targeted outreach methods (such as sending mail to the address) to notify the residents of the dangers of lead-based paint and bring them to http://gettheleadoutgr.org/ for more information about how they can receive assistance. The mailing should also include straightforward information about the health concerns surrounding lead-based paint so that people take the issue seriously. This method of outreach will help the people that need it most. Lastly, we recommend that the City of Grand Rapids uses both the risk model and the cost model to apply for future federal grants. Our models can be used to demonstrate not only the seriousness of the issue in Grand Rapids, but also displays the research behind how the money is being spent. This strategy can help continue the efforts of the *Get the Lead Out!* program and in the long run, increase the safety of Grand Rapids citizens.

Cost Model

Since we are trying to remove lead from as many houses as we possibly can with $4,200,000, we wanted to make sure that the source of outreach we chose would be beneficial to the program and not a waste of the grant money. We started to look into the available means of outreach and the prices associated with each. We found that billboards were quite expensive, it costs the city $0.46 to mail out a letter, $150-$200 for a mass order of brochures, and $0.20 each for a door hanger. As a group, we decided that the billboard was too expensive and would not draw as much interest as the other options, so we decided to not include the cost of a billboard in our cost model. We're hoping that using our predictive model to reach out to the high risk homes, we can notify them about the dangers of lead-based paint and get more applicants.

The census tracts with the highest number of at risk houses are 32, 16, 35, and 37, respectively. In order to understand the scope of the costs and be mindful of the grant amount, we will only discuss costs of these 4 census tracts. We will not discuss the cost of full removal in these tracts, as the cost to do this either by an owned house or rented house is well over the available funding of the grant. Full removal costs are moreover acting as a ceiling limit on price in order to help as many households as possible we will look at outreach and inspection costs. We also will only briefly touch on total encapsulation cost, as doing so costs over a million USD for each of the four aforementioned census tracts.

In census tract 32, we have 1,478 houses in the 6 highest risk clusters. In order to provide outreach for all of these homes, the cost would be $783.34 by mail and $295.60 using door hangers. Either method would be effective at reaching the households, but the door hanger outreach cost does not include labor of putting up the hangers, as we could not find a reliable price estimate for labor. We then estimate the assessment costs: if we do a swab kit on each house the cost will be $44,340, an x-ray inspection by a professional will cost $443,400, and a professional full assessment is $872,020. The choice of evaluation can be determined on a case to case basis, thus allowing for a chance to reduce overall costs. If we wished to encapsulate all houses based on the average cost needed for the paint, it would cost $1,773,600. For full removal, assuming all houses are not rented, is $22,520,286. Assuming every house is rented, full removal will cost $17,697,572. As noted, full removal in this lone census tract is well above our grant value.

Census tract 16 has 1,475 houses in the identified clusters. Outreach by mail will cost $781.75 and door hangers without labor will cost $295. Assessment choices for swab kits, x-ray inspection, and risk assessment cost $44,250, $442,500, and $870,250, respectively. If we wished to do full removals, assuming all owned houses would cost $22,474,575, assuming all rental houses would cost $17,661,650, and encapsulation would be $1,770,000.

Census tract 35 contains 1,359 houses in the set of high risk clusters. Outreach cost by mail would be $720.27, and by door hanger would cost $271.80 without labor. Assessment choices by either swab kits, x-ray inspection, or risk assessment costs $40,770, $407,700, and $801,810, respectively. If we chose to do full removal in this census tract and we assume every house is owned, it costs $20,707,083, if all rented $16,272,666, and if we encapsulate all houses, $1,630,800.

Finally, census tract 37 contains 1,238 houses in the specified clusters. Outreach cost by mailing is $656.14, outreach by door hangers is $247.60 without labor costs. Assessing each house in the tract by swab kit, x-ray, and risk assessor costs $37,140, $371,400, and $730,420, respectively. Full removal in this tract assuming all houses are owned is $18,863,406, assuming all properties are rental properties is $14,823,812, and full encapsulation in this tract will cost $1,485,600.

Keeping in mind the limits of grant funding, there would be significant financial barriers to completely resolve the problem of lead in Grand Rapids. Our evaluation of costs shows that even attempting to remedy one census tract is nearly infeasible. For the most

efficient use of funding, we would recommend outreach to each household through our two defined options of mail or door hangers. Also, we would recommend an application process to determine which households are in dire need for lead remediation. After approval of an application, we advise sending out one of the options for risk assessment. Only after the risk assessment is complete would the option of interim controls or full abatement be implemented. This would save a significant amount of money as we would not be testing and assessing every possible house.

Acknowledgments

We would like to recognize the invaluable assistance provided by our liaisons at the City of Grand Rapids, Becky Jo Glover, Jon Oeverman, Alex Melton, and Zac Thiel. Their professional guidance and support has challenged and stretched us to create a final project we are proud of. Additionally, we would like to show our gratitude to Dr. David Austin who consistently provided fascinating insight and expertise that was essential for the execution of our project. We could not have done this without him.