Safe at Home? Exploring the Relationship Between Housing Environment and Children's Lead Exposure

Nicole Sosa & Seoeun Hong
nls406 & sh6348
New York University

Abstract

This data analysis project aimed to investigate the potential relationship between housing conditions and elevated blood lead levels in children, as well as to explore the role of economic status in this relationship. Using Hadoop Distributed File System (HDFS), Spark, and Hive, we analyzed datasets on housing conditions, children's blood lead levels, and economic status for various counties in New York State. We found a weak positive correlation between annual average salary and housing conditions, suggesting that as economic status increases, the number of housing complaints associated with lead exposure also tends to increase in that county. However, we did not find any significant correlation between housing conditions and both ranges of blood lead levels in children, or between housing conditions and economics, based on our CORR() function results. Nevertheless, our visual diagrams revealed some patterns and relationships worth noting. We observed a relationship between higher numbers of housing condition complaints related to lead issues and higher children's blood lead levels in some counties, such as Erie.

1 Introduction

Lead exposure in children has been an ongoing public health issue, still happening in more developed countries like the United States (Crabbe et al., 2022). Lead exposure can stem from a variety of sources such as lead contaminated water and soil, but lead painting and lead dust in housing are the most common sources of exposure. This is significant because lead is a stable element which when released into the environment, can stay there for a long time because it's not broken down easily. Having this in the environment can be consequential to young children because lead exposure can cause them serious health effects because their bodies are still developing and don’t have the strong developed biological system like adults. Moreover, lead exposure can affect the neurological and motor functions of children, which can result in deficits in their learning and physical skills to name a few (Abelsohn & Sanborn, 2010).

Our data analysis project aims to investigate the relationship between housing conditions and elevated blood lead levels in children, as well as the potential impact of economic status. Specifically, we seek to determine the strength of the association between housing conditions and elevated blood lead levels and to examine whether there are any disparities based on economic status. Understanding the specific sources and factors that can contribute to lead poisoning in children can inform targeted prevention and intervention strategies. For instance, if our findings suggest that economic status is a significant factor in lead exposure, this could inform the development of lead poisoning prevention programs to target individuals of lower income and reduce lead poisoning among them.

To conduct this data analysis project, we first ingested all of our datasets on Hadoop Distributed File System (HDFS), which provided fault tolerance and easy access to data for big data tools like Hive and Spark. Next we utilized Spark for our initial profiling and cleaning of our raw datasets. Then, we used Hive for an additional profiling of the dataset to ensure that the datasets were as cleaned as possible. Finally for the analysis, we used Hive on each of our datasets and the combined dataset. Hive was efficient to use for our analysis because we could use HiveQl, which allowed us to easily query against the dataset and gain valuable insights. This can be visually seen below in figure 1:

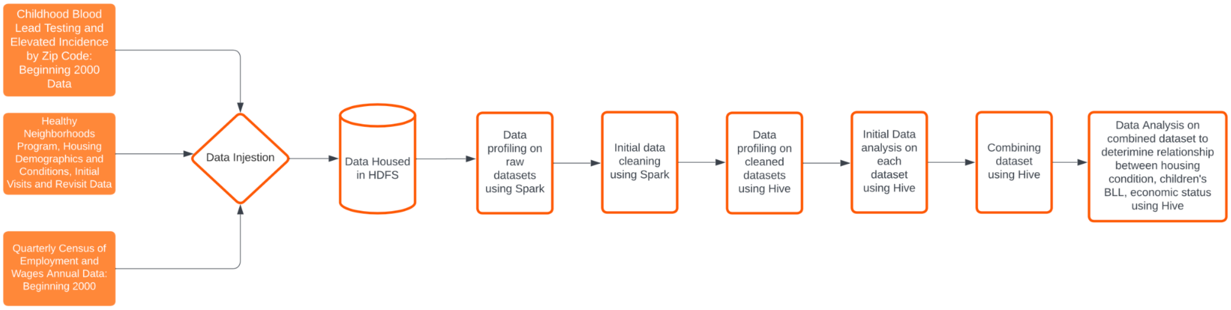

Childhood Blood Lead Testing and Elevated Incidence by Zip Code: Beginning 2000 Data
Healthy Neighborhoods Program, Housing Demographics and Conditions, Initial Visits and Revisit Data
Quarterly Census of Employment and Wages Annual Data: Beginning 2000
Data Injestion
Data Housed in HDFS
Data profiling on raw datasets using Spark
Initial data cleaning using Spark
Data profiling on cleaned datasets using Hive
Initial Data analysis on each dataset using Hive
Combining dataset using Hive
Data Analysis on combined dataset to deterimine relationship between housing condition, children's BLL, economic status using Hive

Figure 1: Data Flow Diagram

Prior to conducting this study, our understanding of the link between lead-related environmental factors and children's blood lead levels was limited to a superficial level, with many questions remaining about which factors have the greatest impact, why this was the case, and how the issue could be effectively addressed. Through our data analysis project, we were able to shed light on some of these questions, uncovering a weak connection between housing conditions and children's blood lead levels. Although our study did not find a significant link between children's blood lead levels and socioeconomic status, it is important to note that other studies, such as studies conducted by Haley and Talbot, have shown that children living in low-income communities with a lower proportion of high school graduates are at a higher risk of lead poisoning (Haley, V. B., & Talbot, T. O., 2004). As such, targeted actions, such as increasing public awareness, campaigning, and implementing policies for the replacement of older homes, are needed in these low-income communities to reduce the risk of lead poisoning and improve the health outcomes of children.

2 Motivation

Although there has been an increase in lead poisoning prevention programs and efforts to increase awareness of lead exposure, lead exposure continues to be a problem and children less than six year olds are the most vulnerable to it. Unfortunately, lead poisoning can be challenging to detect and typically the symptoms become more apparent once children have high levels of lead in their system.With more children spending time indoors than outside, it is crucial to investigate the impact of housing conditions on children's blood lead levels, given that housing environments are the most common source of lead exposure (Cohen 2023).

Furthermore, it is important to understand the potential socioeconomic factors that may contribute to the prevalence of lead exposure in certain housing conditions. Low-income families may not have the financial means to maintain their homes or live in safer conditions and lack access to resources or information regarding lead exposure.

3 Related Work

3.1 Relationship between Children Blood Lead Level and Housing Environment

Many studies have shown the close relationship between children's blood lead level and housing environment.

Kim et al. observed a dose-response trend between children's blood lead levels and the age of housing in Jefferson County, KY, with older houses associated with higher lead levels. Moreover, the authors found that within the older housings, houses with lower values were more likely to pose a greater risk of lead exposure to children (Kim, 2002).

Leighton et al. investigated the impact of lead-based paint hazard remediation on children's blood lead levels in New York City. The authors figured out that children living in homes with earlier remediation had greater declines in blood lead levels compared to children in homes with later remediation (Leighton, 2003).

3.2 Relationship between Children Blood Lead Level and Economic Status

Numerous studies have linked high levels of lead in children's blood to lead-based paint in housing, as well as socioeconomic status, because people with lower socioeconomic status typically tend to live in older and less valuable houses.

Kim et al. also highlighted that children living not only in older houses but also from lower-income families are at the greatest risk of being exposed to lead-painting environments (Kim, 2002).

Lisa M. Nicholson, et al. also found that children from lower-income families are eight times more likely to suffer from lead poisoning compared to those from upper-income families based on CDC data. Moreover, children who have been adopted from developing countries are at a higher risk of lead exposure than those born in the United States (Lisa M. Nicholson, 2010).

4 Datasets

4.1 Childhood Blood Lead Testing and Elevated Incidence by Zip Code: Beginning 2000

We obtained the most recent version of Childhood Blood Lead Testing and Elevated Incidence by Zip Code: Beginning 2000 dataset, last updated on March 8, 2022, in a CSV format from Open Data NY website. This dataset was provided by the New York State Department of Health and it includes information since the year 2000 on the number of children tested for lead who live in the New York state zip code. Furthermore, this dataset provides information on the number of children found with elevated concentrations of lead in their blood for the first time, as well as the number of children who did not have elevated concentration of lead in their blood.

4.2 Healthy Neighborhoods Program, Housing Demographics and Conditions, Initial Visits and Revisits

We obtained the most recent version of the Healthy Neighborhoods Program dataset, provided by the New York State Department of Health and obtained in CSV format from the Open Data NY website, contains information on the housing conditions of each county in New York State. The data was last updated on May 20, 2019 and is based on visits to dwellings by county health departments. The dataset includes details on dwelling characteristics (i.e. age) and the presence or absence of 34 different housing conditions like (i.e. lead paint hazards.)

4.3 Quarterly Census of Employment and Wages Annual Data: Beginning 2000

We obtained the most recent version of Quarterly Census of Employment and Wages Annual Data: Beginning 2000 dataset, provided by New York State Department of Labor and obtained in CSV format from the Open Data NY website. This dataset provides information on employment and wage data from most nonfarm employers in New York State, covering private-sector and government employees insured under the state's unemployment insurance law, while excluding certain worker categories such as agricultural workers. The dataset, which is based on quarterly tax reports submitted by all employers subject to the unemployment insurance law, was last updated on November 3, 2022.

4.4 Attribute Selection on Each Dataset

We utilized three datasets to explore the correlation between children's elevated blood lead levels and their housing environment, while also taking into account the potential economic fact of whether children with lower income are more likely to have elevated blood lead levels when living in certain housing conditions. We selected specific attributes from each dataset to join tables and gain useful insights. Specifically, for each dataset we selected at least one attribute that contains the county name so we could join on that. Table 1 shows the attributes and respective type we selected from the Childhood Blood Lead Testing and Elevated Incidence by Zip Code: Beginning 2000 dataset. Table 2 shows the attributes and respective types we selected from the Healthy Neighborhoods Program, Housing Demographics and Conditions, Initial Visits and Revisits. Table 3 shows the attributes and respective type we selected from the Quarterly Census of Employment and Wages Annual Data: Beginning 2000.

Attributes	**Type**
County	Plain Text
Year	Number
Less than 5 mcg/dL	Number
5-10 mcg/dL	Number

Table 1: Selected attributes and types of Childhood Blood Lead Testing and Elevated Incidence by Zip Code: Beginning 2000 dataset

Attributes	**Type**
Funding Cycle	Plain Text
County Name	Plain Text
Variable	Plain Text
Frequency Count	Number
Percent of Total Frequency	Number

Table 2: Selected attributes and types of Healthy Neighborhoods Program, Housing Demographics and Conditions, Initial Visits and Revisits

Field Name	**Data Type**
Area	Plain Text
Year	Plain Text
Annual Average Salary	Number

Table 3: Selected attributes and types of Quarterly Census of Employment and Wages Annual Data: Beginning 2000

5 Analytic Stages

5.1 Ingestion

We utilized three different datasets which were publicly available from Open Data NY. These were provided in a CSV format, which we downloaded from the Open Data NY website. Then we uploaded all of our datasets to Dataproc from our local machine and then transferred them to a directory in the Hadoop Distributed File System (HDFS).

Note to make it easier to refer to the three datasets we used throughout the paper, we have given them aliases. The first dataset, "Healthy Neighborhoods Program, Housing Demographics and Conditions, Initial Visits and Revisits" dataset will be referred to as the "Housing Conditions" dataset. The "Childhood Blood Lead Testing and Elevated Incidence by Zip Code: Beginning 2000" dataset will be referred to as the "Children Blood Lead Level (BLL)" dataset. Finally, the "Quarterly Census of Employment and Wages Annual Data: Beginning 2000," will be referred to as the "Annual Wages" dataset.

5.2 Initial Profiling on Raw Datasets

To perform our initial data profiling on these raw datasets, we used Spark Scala. Using the select and distinct functions, we were able to view all the data values and unique values per column that we were interested in analyzing. This allowed us to gain a better understanding of the data and identify any potential issues or inconsistencies.

For instance, we found that the “Area” column in the Annual Wages dataset contained all the county names with the word county in them (i.e. Albany County) while all the other dataset contained only the name (i.e Albany). Moreover, we found that it would be best if we normalized all of our textual data to be in uppercase format so there won’t be any inconsistencies when joining. Lastly, we realized that the Housing Conditions dataset contained some unusual values in the “Variable”column not related to housing like“Race” and “Hispanic.”

5.3 Cleaning of Raw Datasets

Based on the data profiling result, we cleaned our datasets using Scala. We loaded each CSV file into a DataFrame and filtered out any unnecessary data. We excluded null values and converted the 'county' variable to uppercase for normalization, so that we could use this column when joining three different datasets. And then, we noticed that the Housing Conditions dataset only contains data for 18 counties between the years 2006 and 2019. Therefore, we extracted the data that satisfied these limitations from our other two datasets, Children Blood Lead Level and Annual Wages.

5.3.1 Children Blood Lead Level

We selected only four columns from our dataset: County, Year, Less than 5 mcg/dL, and 5-10 mcg/dL. We chose to include only the 'County' column for location information, and we selected 'Less than 5 mcg/dL' and '5-10 mcg/dL' because these values are the standard way to divide low and high blood lead levels. Moreover, the other columns contained many null values and some values appeared to be outliers. Finally, we kept the 'Year' column to extract data only from 2006 to 2019.

5.3.2 Housing Conditions

We selected only five columns from our dataset: Funding Cycle, County Name, Variable, Frequency Count, and Percent of Total Frequency. We included the 'County Name' column for location information. We selected the 'Variable' column because it contains the specific reasons for housing condition complaints. We noticed that among all values, four variable reasons - 'Chemical smell indoors', 'Deteriorated paint inside', 'Age of housing', and 'Deteriorated paint outside' - are related to lead housing conditions. Therefore, we filtered out the rows that only had variable values among these four variable reasons. We also extracted the 'Frequency Count' and 'Percent of Total Frequency' columns to count the number of housing condition complaints.

5.3.3 Annual Wages

To examine the economic status of different counties, we selected only three columns: County, Year, and Annual Average Salary. We included the 'County' and 'Year' columns for the same reasons as the other datasets. We also selected the 'Annual Average Salary' column to use this amount as an indicator of the economic status of each county.

5.4 Profiling of Cleaned Datasets

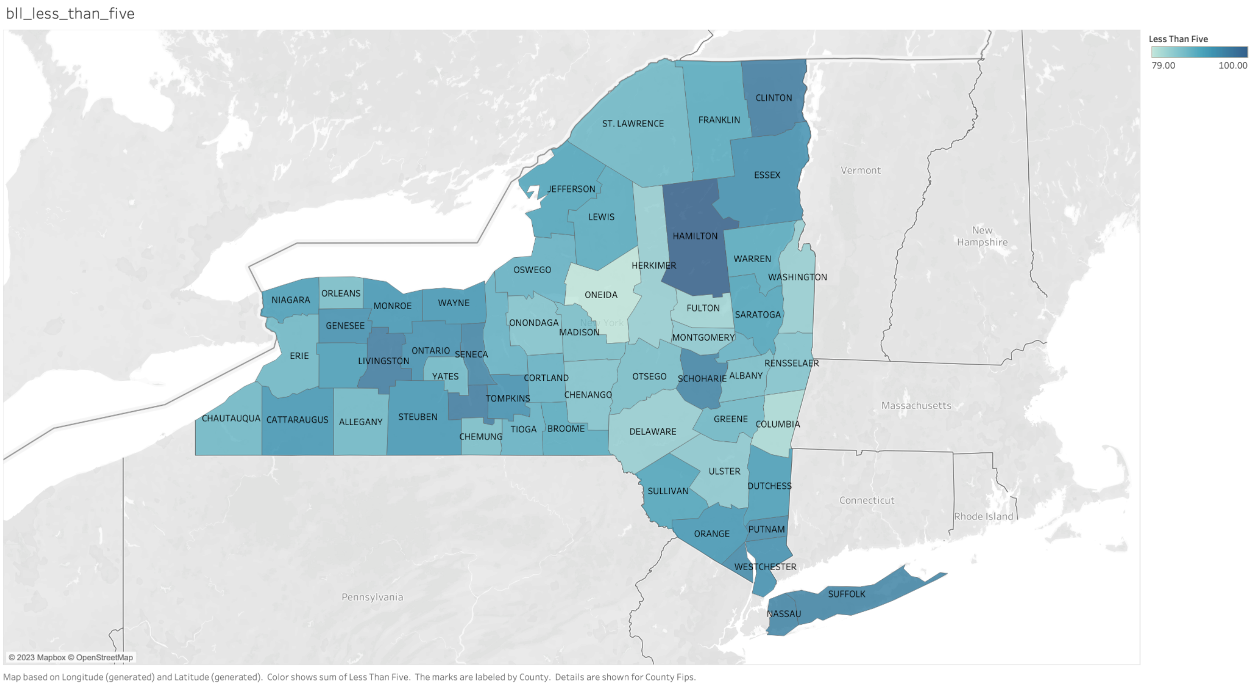
After cleaning our datasets, we conducted data profiling on them. For datasets that contained numeric data, such as children's blood lead levels and annual wages, we performed statistical tests to obtain metrics such as finding the percentile, minimum, maximum, and average values for each county and for the dataset as a whole. For datasets that contained categorical data, such as housing conditions, we performed different types of analysis rather than statistical tests, such as examining the frequency and uniqueness of the values.

5.5 Analytic of Individual and Merged Datasets

After performing data cleaning using Scala, we stored all the cleaned datasets into separate external Hive tables and ran HiveQL queries to perform analysis on them. In addition, as shown in Figure 6, we used an "inner join" operation to combine these external tables based on their location column (i.e. county or area). However, during the join process, we did lose a considerable amount of information, since the housing conditions dataset only included data from a limited number of counties.

5.5.1 Children Blood Lead Level Less than 5 mcg/dL (Low Blood Lead Level)

Instead of counting the actual number of children with low blood lead levels in each county, we calculated the relative percentage by dividing the number of children with low blood lead levels by the total number of children in that county. This method was used to account for differences in population between counties, which may affect our results. In figure 2, we can visually see the counties with the lowest and highest number of children with low blood lead levels as shown below:


bll_less_than_five

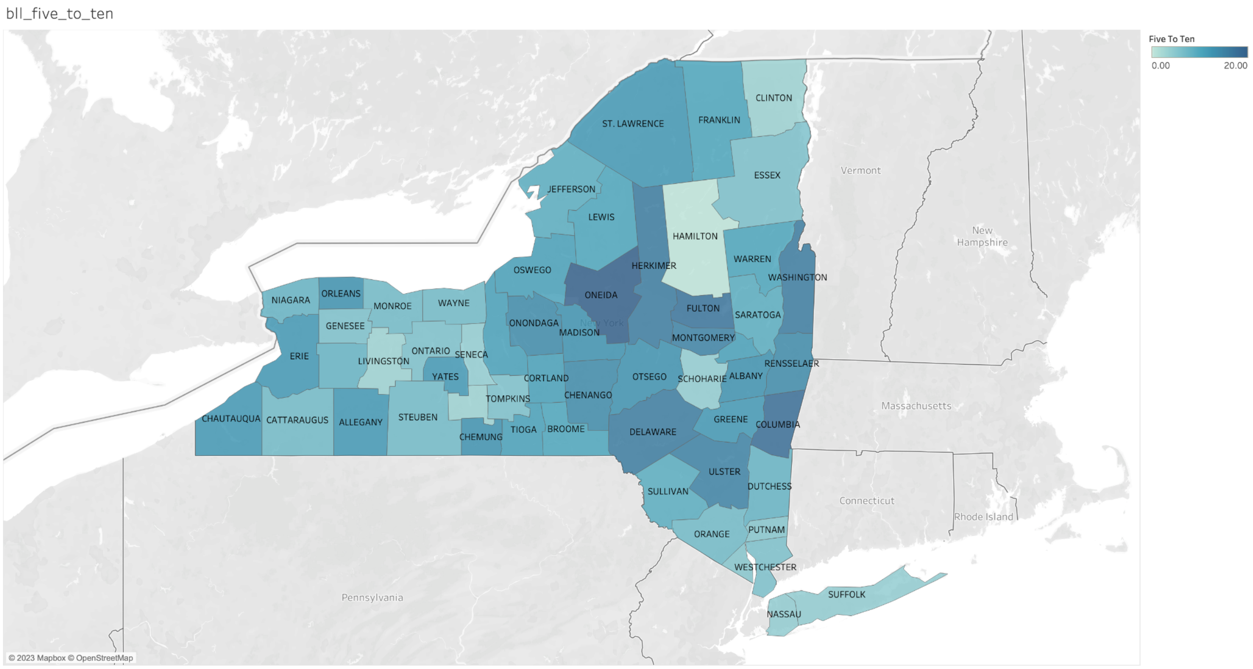
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Less Than Five. The marks are labeled by County. Details are shown for County Fips.

Figure 2: visualization of the number of children with low blood lead levels in each county

Based on Figure 2, we observed a weak trend, but it still appeared that children who live closer to the New York City area tended to have lower blood lead levels compared to those who live upstate. Hamilton County had the lowest average blood lead level among the counties analyzed. However, counties located near the New York City area, such as Westchester, Rockland, Nassau, and Suffolk, tended to show lower children blood lead levels among.

5.5.2 Children Blood Lead Level 5 to 10 mcg/dL (Elevated Blood Lead Level)

To address differences in population size across counties, we opted to use a proportional approach rather than simply counting the number of children with elevated blood lead levels. Specifically, we calculated the percentage of children affected in each county by dividing the number of children with elevated blood lead levels by the total number of children in that county. This method enabled us to make more accurate comparisons of elevated blood lead levels between counties and to identify any underlying patterns or trends that may exist. In figure 3, we can visually see the counties with the lowest and highest number of children with elevated blood lead levels as shown below:



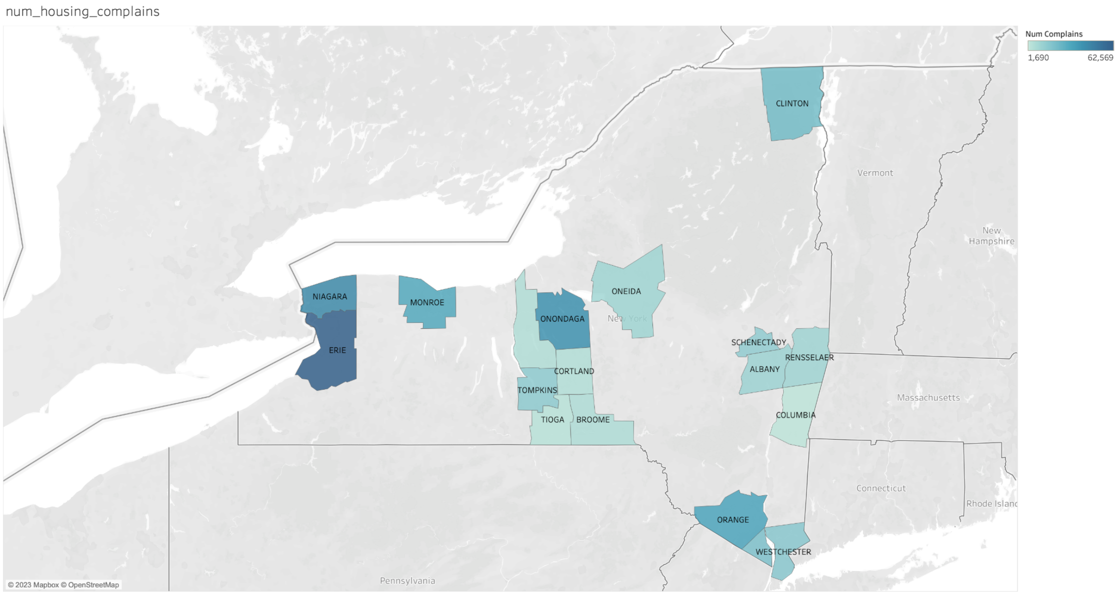
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Five To Ten. The marks are labeled by County. Details are shown for County Fips.

Figure 3: visualization of the number of children with elevated blood lead levels in each county

Based on this visualization, we noticed that Oneida, Fulton, and Columbia were the counties with the highest number of children with elevated blood lead levels. In contrast, Hamilton, Schoharie, Livingston, and Seneca were the counties with the lower number of children with elevated blood lead levels. Additionally, we observed that children's blood lead levels tended to increase as we moved farther from New York State. However, there was an interesting pattern where the levels decreased as we moved west, only to increase again afterwards.

5.5.3 Housing Conditions

In this dataset, each county had a frequency count of the number of houses that had the same complaint. We wanted to determine the total count of the frequency of housing complaints associated with conditions that can lead to lead exposure in each county. We accomplished this by using group aggregation with the sum() function. In figure 4 below, we can see the counties with the highest number of complaints associated with conditions that can lead to lead exposure.



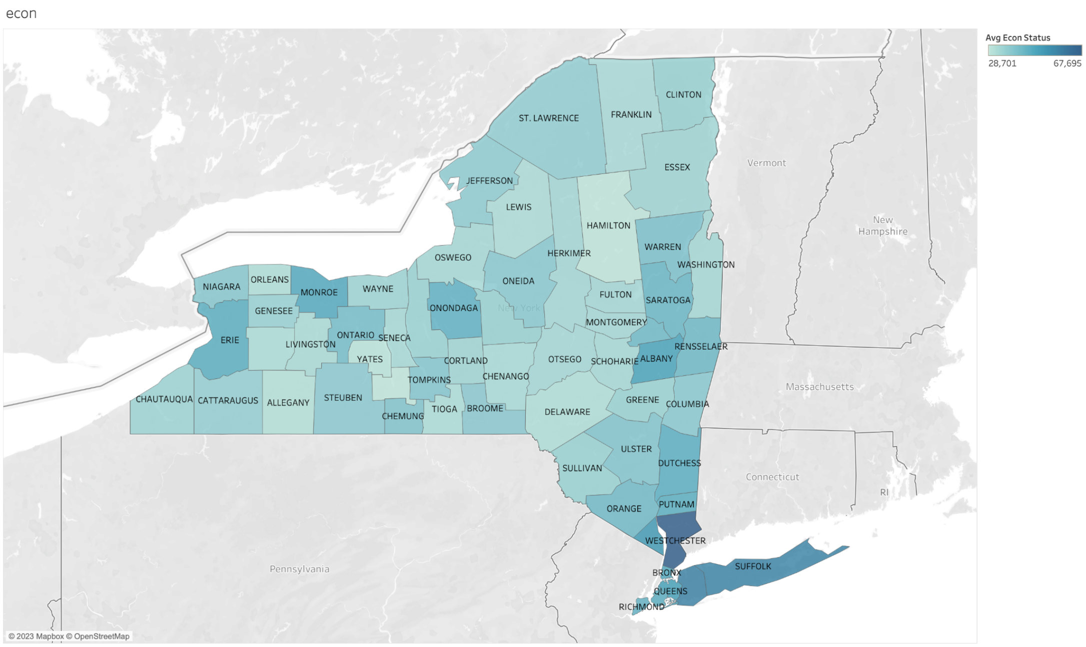
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Num Complains. The marks are labeled by County. Details are shown for County Fips.

Figure 4: shows the frequency counts of housing complaints related to conditions that can lead to lead exposure in each county

Based on this visualization, we noticed that Erie had the highest number of complaints that were associated with lead exposure conditions like Deteriorated paint inside and Deteriorated paint outside. Following Erie, Onondaga and Niagara were the counties with the highest number of complaints as well.

5.5.4 Annual Wages

We grouped the annual average salary by county and calculated the average for each county, as different regions within each county have their own unique annual average salary.



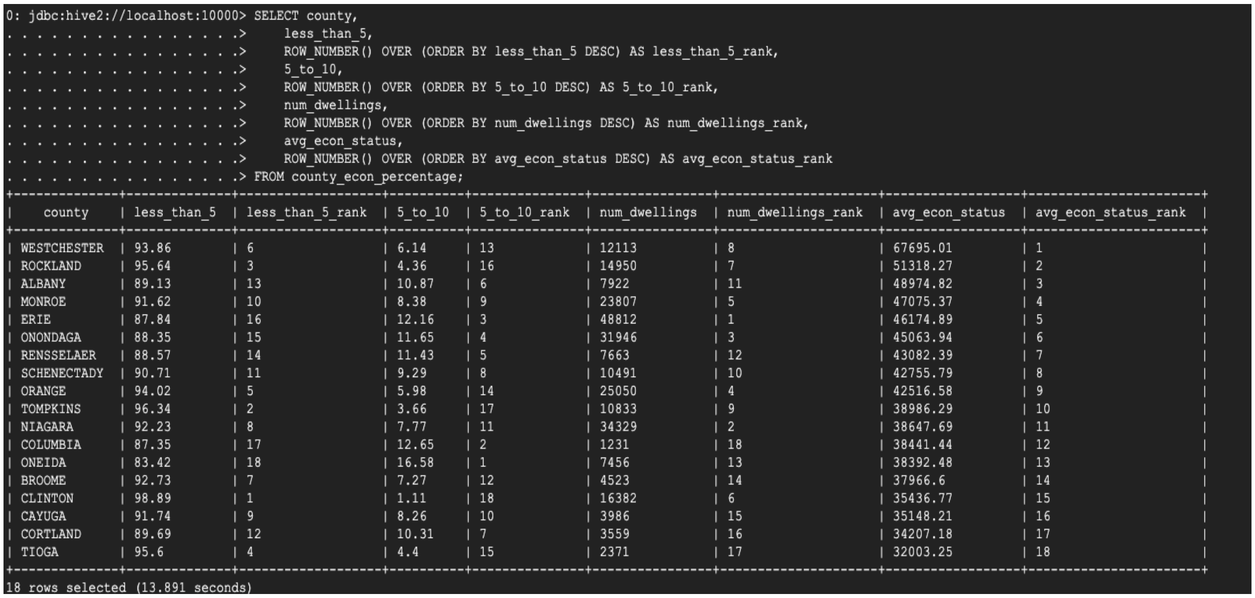
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Avg Econ Status. The marks are labeled by County. Details are shown for County Fips.

Figure 5: Visualization of the average annual wages for each New York State county

The data shows that people who live closer to the New York City Area tend to have a higher economic status compared to those living in upstate. We observed a weak but somewhat relationship between children's blood lead levels and economic status - when the economic status is higher, the children's blood lead levels tend to be lower, and vice versa.

5.5.5 Merged Datasets

To show the exact numeric order, we utilized window functions and the RANK() function to generate a table with a unique rank assigned to each county based on their values for each column. This allowed us to avoid rough estimations. In addition to this, this provides a precise ranking order for the data and allows us to do statistical analysis based on their ranking.



```
0: jdbc:hive2://localhost:10000> SELECT county,
. . . . . . . . . . . . . . . .>     less_than_5,
. . . . . . . . . . . . . . . .>     ROW_NUMBER() OVER (ORDER BY less_than_5 DESC) AS less_than_5_rank,
. . . . . . . . . . . . . . . .>     5_to_10,
. . . . . . . . . . . . . . . .>     ROW_NUMBER() OVER (ORDER BY 5_to_10 DESC) AS 5_to_10_rank,
. . . . . . . . . . . . . . . .>     num_dwellings,
. . . . . . . . . . . . . . . .>     ROW_NUMBER() OVER (ORDER BY num_dwellings DESC) AS num_dwellings_rank,
. . . . . . . . . . . . . . . .>     avg_econ_status,
. . . . . . . . . . . . . . . .>     ROW_NUMBER() OVER (ORDER BY avg_econ_status DESC) AS avg_econ_status_rank
. . . . . . . . . . . . . . . .> FROM county_econ_percentage;
+--------------+--------------+-------------------+----------+---------------+----------------+---------------------+------------------+-----------------------+
|    county    | less_than_5  | less_than_5_rank  | 5_to_10  | 5_to_10_rank  | num_dwellings  | num_dwellings_rank  | avg_econ_status  | avg_econ_status_rank  |
+--------------+--------------+-------------------+----------+---------------+----------------+---------------------+------------------+-----------------------+
| WESTCHESTER  | 93.86        | 6                 | 6.14     | 13            | 12113          | 8                   | 67695.01         | 1                     |
| ROCKLAND     | 95.64        | 3                 | 4.36     | 16            | 14950          | 7                   | 51318.27         | 2                     |
| ALBANY       | 89.13        | 13                | 10.87    | 6             | 7922           | 11                  | 48974.82         | 3                     |
| MONROE       | 91.62        | 10                | 8.38     | 9             | 23807          | 5                   | 47075.37         | 4                     |
| ERIE         | 87.84        | 16                | 12.16    | 3             | 48812          | 1                   | 46174.89         | 5                     |
| ONONDAGA     | 88.35        | 15                | 11.65    | 4             | 31946          | 3                   | 45063.94         | 6                     |
| RENSSELAER   | 88.57        | 14                | 11.43    | 5             | 7663           | 12                  | 43082.39         | 7                     |
| SCHENECTADY  | 90.71        | 11                | 9.29     | 8             | 10491          | 10                  | 42755.79         | 8                     |
| ORANGE       | 94.02        | 5                 | 5.98     | 14            | 25050          | 4                   | 42516.58         | 9                     |
| TOMPKINS     | 96.34        | 2                 | 3.66     | 17            | 10833          | 9                   | 38986.29         | 10                    |
| NIAGARA      | 92.23        | 8                 | 7.77     | 11            | 34329          | 2                   | 38647.69         | 11                    |
| COLUMBIA     | 87.35        | 17                | 12.65    | 2             | 1231           | 18                  | 38441.44         | 12                    |
| ONEIDA       | 83.42        | 18                | 16.58    | 1             | 7456           | 13                  | 38392.48         | 13                    |
| BROOME       | 92.73        | 7                 | 7.27     | 12            | 4523           | 14                  | 37966.6          | 14                    |
| CLINTON      | 98.89        | 1                 | 1.11     | 18            | 16382          | 6                   | 35436.77         | 15                    |
| CAYUGA       | 91.74        | 9                 | 8.26     | 10            | 3986           | 15                  | 35148.21         | 16                    |
| CORTLAND     | 89.69        | 12                | 10.31    | 7             | 3559           | 16                  | 34207.18         | 17                    |
| TIOGA        | 95.6         | 4                 | 4.4      | 15            | 2371           | 17                  | 32003.25         | 18                    |
+--------------+--------------+-------------------+----------+---------------+----------------+---------------------+------------------+-----------------------+
18 rows selected (13.891 seconds)
```

Figure 6: Joined table on location(i.e. county, area) column for merged datasets

Based on the table, we can see that the top three counties with the lowest children's blood lead levels are Clinton, Tompkins, and Rockland, while the bottom three counties with the highest children's blood lead levels are Oneida, Columbia, and Erie. Additionally, we can see that the top three counties with the highest number of housing condition complaints related to lead issues are Erie, Niagara, and Onondaga, while the bottom three counties with the lowest number of housing condition complaints are Cortland, Tioga, and Oneida.

It is important to note that not all counties exhibit the same trend, but there does seem to be a correlation between higher numbers of housing condition complaints related to lead issues and higher children's blood lead levels in some counties, such as Erie.

Westchester, Rockland, and Albany are among the top counties with higher economic status, while Cayuga, Cortland, and Tioga are among the counties with lower economic status. However, there appears to be no clear relationship between economic status and the number of housing complaints related to lead issues.

We performed correlation tests using the CORR() function in Hive on our combined dataset. Our analysis revealed a negligible correlation of 0.047 between housing conditions and children with blood lead levels in the range of 5 to 10 mcg/dL, as well as a negligible correlation of -0.047 between housing conditions and children with blood lead levels less than 5 mcg/dL. Furthermore, we found negligible correlations of -0.000877 between children with blood lead levels in both ranges and average annual salary (represented as the "avg econ status" column in the merged table). However, we did find a weak positive correlation of 0.25 between annual average salary and housing conditions, which suggests that as economic status increases the number of housing complaints associated with lead exposure also tends to increase in that county.

7 Conclusion

This data analysis project aimed to investigate the potential relationship between housing conditions and elevated blood lead levels in children, as well as to explore the role of economic status in this relationship. While using Hive's CORR() function to calculate the correlation coefficient, we did not find any significant correlation between housing conditions and both ranges of blood lead levels in children, or between housing conditions and economics. However, our visual diagrams (figures 2 - 6) revealed some patterns and relationships worth noting.

Based on our observations from figures 2, 3, and 5, counties located near the New York City area tended to have higher economic status and a larger proportion of children with low blood lead levels compared to elevated blood lead levels. This finding suggests that economic status could potentially play a role in a child's risk of lead exposure. Additionally, when examining the county rankings in figure 6, we found that Erie County was among the top three counties with the highest number of children with elevated blood lead levels, as well as one of the top three counties with the most housing condition complaints related to lead issues. This observation indicates a potential relationship between lead exposure and poor housing conditions.

However, it is important to note that our housing conditions dataset did not include data for all counties in New York State, leading to a loss of data during the join process. This could have affected our results and led to the low correlation and negligible correlations obtained. In future research, it would be beneficial to explore alternative housing condition datasets that include data for all counties to retain complete data. While the Children Blood Lead Levels and Annual Average Salary datasets did have data for all counties in New York State, we still lost data when joining all three datasets on the county column.

Acknowledgments

We thank NYU High Performance Computing (HPC) for their support and Prof. Malavelt for giving us guidance on this project.

References

Abelsohn AR, Sanborn M. Lead and children: clinical management for family physicians. Can Fam Physician. 2010 Jun;56(6):531-5. PMID: 20547517; PMCID: PMC2902938.

Cohen, D. (2023, April 14). *Why kids need to spend time in nature*. Child Mind Institute. Retrieved May 4, 2023, from https://childmind.org/article/why-kids-need-to-spend-time-in-nature/#:~:text=These%20days%2C%20kids%20spend%20much,in%20front%20of%20a%20screen

Crabbe, H., Verlander, N. Q., Iqbal, N., Close, R., White, G., Leonardi, G. S., & Busby, A. (2022). ‘As safe as houses; the risk of childhood lead exposure from housing in England and implications for public health.’ *BMC Public Health*, *22*(1). https://doi.org/10.1186/s12889-022-14350-y

Haley, V. B., & Talbot, T. O. (2004). Geographic analysis of blood lead levels in New York State children born 1994-1997. *Environmental health perspectives*, *112*(15), 1577–1582. https://doi.org/10.1289/ehp.7053

Kim, D. Y., Staley, F., Curtis, G., & Buchanan, S. (2002). Relation between housing age, housing value, and childhood blood lead levels in children in Jefferson County, Ky. *American journal of public health*, *92*(5), 769–772. https://doi.org/10.2105/ajph.92.5.769

Leighton, J., Klitzman, S., Sedlar, S., Matte, T., & Cohen, N. L. (2003). The effect of lead-based paint hazard remediation on blood lead levels of lead poisoned children in New York City. *Environmental Research*, *92*(3), 182–190. https://doi.org/10.1016/s0013-9351(03)00036-7

Lisa M. Nicholson, Kent P. Schwirian, & Patricia M. Schwirian. (2010). Childhood Lead Poisoning Laws in New York City: Environment, Politics and Social Action. *Children, Youth and Environments*, *20*(1), 178–199. http://www.jstor.org/stable/10.7721/chilyoutenvi.20.1.0178