Leika Ray Joseph

Capstone Project

DATA 205

05/12/2025

Lori Perine

**Crime, Home Value, and School Exposure: Insights from a Spatial Analysis of Montgomery County**

Understanding the potential relationships between crime, housing, and education is crucial for improving community well-being. In Montgomery County, Maryland, these factors may influence one another in complex ways, and exploring these connections can provide valuable insights for residents, local leaders, and policymakers. This project investigates crime trends in Montgomery County and examines how these trends might be related to property values and the performance of public schools. The analysis uses several important datasets to explore these connections. The Crime Dataset from Data Montgomery includes over 437,000 reported crimes in the county since 2016, providing details about crime types, locations, and other relevant information. While there are some challenges, such as missing data and incorrect locations, this dataset is essential for understanding crime trends. The Census Dataset from the U.S. Census Bureau API provides demographic and economic information for cities in Montgomery County, including the median value of homes, which helps us look at whether areas with higher crime rates have lower property values. Finally, the Public School Dataset from Data Montgomery offers basic details about around 200 public schools, such as school locations, types, and contact information. Although it doesn't include data on school performance or student numbers, this dataset is useful for mapping school locations in relation to crime areas and looking at how local crime might affect schools.

The primary objectives of this project are to examine crime patterns in Montgomery County and understand how they have shifted over time and across different geographic areas. A key focus is to investigate whether variations in crime levels are related to differences in median home values across cities, with the aim of determining if communities with higher crime levels tend to have lower property values. Additionally, the project explores how public schools may be influenced by the surrounding crime environment, seeking to identify potential connections between neighborhood safety and educational settings. By combining data cleaning, statistical analysis, and visualization techniques, this project aims to provide a comprehensive picture of how crime intersects with important social and economic factors such as housing markets and education, offering insights that may support future research or inform local policy discussions.

For this project, I used **OpenRefine**, **RStudio**, and **Python** as the main tools. I started with OpenRefine to clean the datasets by fixing missing values and correcting errors like misspelled city names, which made the data ready for analysis. Next, I used RStudio for more data cleaning and to create visualizations such as bar plots, alluvial diagrams, and maps. I also used RStudio to look at the relationship between crime numbers and median home values across cities. After preparing the data in R, I exported the cleaned datasets and used Python for additional analysis. In Python, I ran t-tests to check how median home values are related to different types of crime and created interactive maps and visualizations to show the geographic spread of schools and crime patterns. The main approaches I used in this project included looking at trends over time, making correlation plots to explore links between crime rates and home values, and using statistical tests to understand how crime may affect property values.

The data cleaning and pre-processing for this project involved several important steps to ensure the datasets were accurate and ready for analysis. For the Crime Dataset, the first step was to use OpenRefine to fix misspellings in city names, ensuring consistency across all datasets. I also standardized the column names by converting them to lowercase and removing spaces, making them easier to work with. The dispatchdate/time column was reformatted into a proper Date format, and new columns for the month and year were added to help analyze crime trends. Reports from 2025 were removed as they weren’t relevant for this project, and the City column was cleaned by eliminating incorrect locations and cities not located in Montgomery County. Some entries had numbers instead of city names, so I used a list of correct city names to filter out errors and ensured the text was lowercase and free of extra spaces. For the Census Dataset, I used an API to gather data on the median home values in cities within Montgomery County. I cleaned the data by making column names lowercase, removing spaces, and ensuring the city names matched those in the Crime Dataset by removing words like "town" or "city." Any cities not part of Montgomery County was removed, and unusual city names were corrected. Finally, the School Dataset required cleaning as well. I ensured that column names were consistent, removed rows with missing or incorrect geographic data, and converted the school data into a spatial format so it could be analyzed alongside the crime data.

Among the cities in Montgomery County, Silver Spring reports the highest number of crimes, followed by Wheaton, Montgomery Village, Bethesda, Rockville, Germantown, and Takoma Park. These areas appear to be crime hotspots within the county, raising important questions about what factors contribute to their higher crime levels. Are these patterns related to population density, economic conditions, urban infrastructure, or other social factors? Understanding the geographic distribution of crime is a key step toward examining how it may connect to other variables in this analysis, such as property values and the safety of nearby schools.

A graph of a crime distribution

AI-generated content may be incorrect.

The cleaned crime dataset includes over 430,000 reported incidents in Montgomery County from 2016 to 2024. Crime records include variables such as crime type, report date/time, location, latitude and longitude, reporting agency, dispatched time and more. The most commonly reported crimes include all other larceny, destruction or damage of property, simple assault, and theft from motor vehicles.

The median home value dataset includes estimates for cities across the county, ranging from approximately $260,000 to over $1 million, with Bethesda and Chevy Chase among the highest. The public school dataset contains basic information on nearly 200 schools, including school type (elementary, middle, or high school), address, and location coordinates. While student performance data was not available, this dataset enabled spatial comparisons between school locations and crime density. Temporally, crime counts are highest between May and July, with lower levels during the winter months. Crime totals dipped in 2020 and 2021, likely due to the impact of the COVID-19 pandemic. Geographically, Silver Spring had the highest number of reported crimes, followed by Wheaton, Montgomery Village, and Bethesda.A graph of crime trends over time by category

AI-generated content may be incorrect.

This line plot shows how crime has been trending over the years in Montgomery County. One thing stands out right away: crime against property is by far the most common type of crime, and it has been steadily increasing over time. So far, there have been 179,020 incidents of property crime, making it the biggest concern in the county. The next largest group is crimes that impact society, with 141,175 incidents, followed by crimes against individuals. The top 10 most frequent crimes are: Theft From Motor Vehicle, Simple Assault, Shoplifting, Destruction/Damage/Vandalism of Property, Drug/Narcotic Violations, Driving Under the Influence, Theft from Building, Motor Vehicle Theft, Identity Theft, and Burglary/Breaking and Entering.

A graph of crime and crime

AI-generated content may be incorrect.

Most property crimes happen in residential parking lots, likely because these areas have less monitoring, especially at night. Cars left with valuables inside become easy targets for theft. Looking at trends over time, we see mixed patterns. Shoplifting has steadily increased each year, while drug and narcotic violations have dropped sharply. Robbery has seen a slight rise, but overall, it remains low compared to other crimes. These shifts may reflect larger forces, such as inflation or the post-pandemic environment, which have changed both social conditions and law enforcement priorities. But what does this mean for the places people care about the most: their homes? Do neighborhoods with more crime see lower home values?

A graph with a red line

AI-generated content may be incorrect.

The analysis shows no meaningful link between overall crime and home values. With a p-value of 0.43 and R-squared of just 2%, home prices explain almost none of the differences in crime counts between cities. Cities with both low and high home values can have either low or moderate crime. However, when focusing on specific violent crimes like Murder, Forcible Fondling, and Statutory Rape the t-tests revealed significant relationships (p < 0.05). This suggests that while general crime has little effect on housing, serious violent crimes may have a stronger impact on property values and neighborhood reputation.

To further understand the potential impact of crime on community institutions, the analysis examined the proximity of public schools to areas with high crime density. Using spatial analysis, each school’s location was compared to nearby reported incidents within a one-mile radius. The results revealed a clear concentration of high-crime exposure among schools in cities like Silver Spring and Takoma Park areas already identified as general crime hotspots. Notably, East Silver Spring Elementary had over 27,000 incidents within its surrounding radius, far exceeding the average. This proximity-based approach offers a deeper understanding of how neighborhood safety concerns might extend into the daily environments of students and educators. It also reinforces the importance of analyzing not only crime rates, but where that crime occurs and who is most affected by it.

A screenshot of a computer

AI-generated content may be incorrect.

To add spatial context to the analysis, a geographic heat map was created to visualize the distribution of crime across Montgomery County. The map displays areas of high and low crime density using a color gradient, with warmer tones indicating higher concentrations of reported incidents. Overlaying the heat map are all public school locations, represented by blue circles. Each circle is sized according to the number of crimes reported within a one-mile radius of that school, offering a visual comparison of how different school communities experience varying levels of exposure to crime. This map reinforces earlier findings from the proximity analysis and highlights how location-specific factors contribute to safety disparities across the county.

A map with many colored dots

AI-generated content may be incorrect.

The spatial analysis of school locations in relation to crime density suggests that not all schools in Montgomery County are situated in equally safe environments. While many schools are located in areas with relatively low levels of reported crime, a significant number particularly in cities like Silver Spring and Takoma Park appear to be within or near high-crime zones. Some schools had tens of thousands of reported incidents within just a one-mile radius. This does not necessarily mean that students are unsafe within school grounds, but it does raise important questions about neighborhood conditions, student commuting routes, and the broader environments in which these schools operate. The findings suggest that while some schools benefit from safer surroundings, others may be more exposed to risks, indicating an opportunity for more targeted planning, investment in community safety, and collaboration between educational and public safety agencies.

This analysis has uncovered a layered and complex picture of crime across Montgomery County one that defies simple explanations. While overall crime counts showed little connection to housing markets, certain violent crimes like murder and forcible fondling revealed statistically significant links to property values, hinting at deeper social and psychological impacts on neighborhoods. At the same time, schools located in areas of high crime density stood out visually and spatially, raising meaningful questions about the environments in which children learn and grow.

So, are our schools located in the safest places possible? The data suggests: not always. While many benefit from safer surroundings, others are positioned within or adjacent to high-crime zones, underscoring the need for more intentional planning and community support. With more data on school performance, socioeconomic conditions, or even resident perceptions of safety a richer, more actionable picture could emerge. I’ve come this far, and here’s a preview of what could be an attraction: an integrated dashboard or interactive planning tool that visualizes the relationships among crime, housing, and education in real time. This would not only support more informed decision-making but also invite deeper engagement from policymakers and the public.

Starting this semester, I faced a lot of uncertainty from not knowing exactly what I wanted to explore, to struggling with choosing a dataset, and later, feeling unsure about the direction my analysis should take. I would like to express my deepest gratitude to Professor Lori Perine, whose continuous support and guidance made all the difference. Professor Perine was consistently available to offer resources, answer questions, and provide thoughtful feedback that helped me improve and stay on track throughout this project. I would also like to thank Professor Celia Evans, Professor Rachel Saidi and professor Abdirisak Mohamed for their encouragement, insightful instruction, and the inspiration they brought to every class. Each of you contributed meaningfully to the development of the skills and confidence that made this project possible. Special thanks also go to the Montgomery County Open Data Portal and the U.S. Census Bureau for providing access to the datasets used in this analysis. I am grateful for the developers of OpenRefine, RStudio, and Python libraries such as pandas, folium, and matplotlib, which allowed me to clean, analyze, and visualize my data effectively. Lastly, I extend appreciation to the broader open-source and public data communities, whose tools and transparency empower students like me to turn data into insight.

References

Data Montgomery. (2025). Crime. <https://data.montgomerycountymd.gov/Public-Safety/Crime/icn6-v9z3/about_data>

Data Montgomery. (2023). Public Schools. <https://data.montgomerycountymd.gov/Education/Public-Schools/772q-4wm8/data_preview> ​

U.S. Census Bureau. (2022). American Community Survey. <https://api.census.gov/data/2022/acs/acs5>​