rijksuniversiteit groningen

Proximity to Public Schools and House Prices: Insights from Allegheny County, Pennsylvania

---

Master Thesis

Rijksuniversiteit Groningen

Abstract

This study utilizes datasets from the Western Pennsylvania Regional Data Center to examine the relationship between house prices and distance from the nearest public school within the same school district in Allegheny County, Pennsylvania. Employing the hedonic model, the analysis reveals a significant negative relationship: on average, each kilometer further from the nearest public school within the same school district corresponds to a 2.9% decrease in house prices, ceteris paribus. Moreover, the impact of proximity varies across different distances, with the initial net effect being positive, peaking at a moderate distance before gradually diminishing with increasing distance. Additionally, this study employs the Chow test to explore spatial heterogeneity in the relationship across urban, suburban, and semi-rural areas, revealing significant variations in the relationship across these contexts. While urban areas experience the most substantial decrease in house prices (14.0%), suburban areas experience more modest decreases (1.2%). However, the results from semi-rural areas are insignificant (0.8%). Overall, this study provides insights into the relationship between house prices and public-school proximity, thereby guiding more informed and strategic decision-making processes.

Keywords: Hedonics, House Prices, Public Schools, School Districts, Spatial Heterogeneity

COLOFON

Title	Proximity to Public Schools and House Prices: Insights from Allegheny County, Pennsylvania
Version	Final
Author	D.R.P. (Daan) Bosman
Supervisor	Dr. M.N. (Michiel) Daams
Assessor	Dr. M. (Mark) van Duijn
E-mail	d.r.p.bosman@student.rug.nl or drpbosman@gmail.com
Date	28-06-2024

1. INTRODUCTION

In the United States, public schools are funded by taxpayer dollars and are therefore free of tuition for students. However, public school choice is generally confined to the school district where households reside, leading households to indirectly pay for access to higher-quality education by competing for housing in school districts served by better-performing schools. As a result, house prices tend to be higher in school districts with top-rated public schools. This phenomenon occurs because households are willing to pay a premium to ensure their children have access to quality education, effectively "bidding up" house prices in desirable school districts (Owusu-Edusei et al., 2007). In addition to seeking quality education, prospective homebuyers also consider the distance to the nearest public school within the school district where they reside when evaluating properties. The assumption is that homes closer to schools are more attractive to families with school-aged children due to shorter commutes and enhanced safety. However, Guntermann & Colwell (1983) shed light on this complex relationship, noting that while proximity to certain activities can positively impact livability and thus house prices, being excessively close may result in either a decrease in livability and consequently house prices or a less significant increase than initially anticipated. This indicates that the relationship between school proximity and house prices is complex, balancing the benefits of accessibility and the potential drawbacks of living too close to certain schools such as increased noise levels, traffic congestion, petty crime, and vandalism

While the influence of schools on house prices is widely acknowledged, existing research has primarily focused on analyzing how school quality affects house prices (Black, 1999; Downes & Zabel, 2002; Figlio & Lucas, 2004; Brasington & Haurin, 2006; Gibbons & Machin, 2008; Machine, 2011). However, there has been significantly less scholarly attention directed towards investigating the impact of school proximity on house prices, particularly in relation to public schools that are located within the same school district as the households reside. This gap is notable given that, in certain large countries, public school options are typically restricted to the districts where households reside. Most studies have analyzed proximity to schools in general contexts, without specifically addressing how proximity to public schools within the immediate school district boundaries impacts house prices. The earliest research on this topic by Emerson (1972), who examine housing in southern Minneapolis, find that house prices rose significantly with increasing distance from the nearest school. However, these findings are viewed as suggestive rather than conclusive, as the statistical significance of the observed relationship was not firmly established. Following this research, Guntermann & Cowell (1983) examine the effect of proximity to primary schools on house prices in Lubbock, Texas, highlighting both positive and negative externalities of proximity to schools. Rosiers et al. (2001) conduct similar research but then in Quebec, Canada, and yield similar results. Owusu-Edusei et al. (2007) provide a more comprehensive study as they control for school type, quality of the school, and distance to other amenities. Their study, encompassing Greenville, South Carolina, expands on previous research by considering a wide array of factors influencing the relationship between school proximity and house prices. Sah et al. (2015) introduce spatial heterogeneity as their findings show that the impact of school proximity on house prices varies depending on whether the area is inland or coastal within San Diego County. Huang & Hess (2018) employe quantile regression for their research in

Oshkosh, Wisconsin. Finally, Metz (2015) and Huang & Dall'Erba (2021) are among the first to specifically examine the effect of proximity to schools with households being confined to certain enrollment zones or school districts. Metz (2015) explores the relationship between distance to assigned public schools and house prices in the Denver Public School District. The study finds that house prices generally decrease with increased distance to schools. Similarly, Huang & Dall'Erba (2021) examine the relationship between proximity to secondary schools and house prices within four school enrollment zones in Auckland, New Zealand, utilizing both the hedonic pricing model and the quantile regression model. Their findings reveal nonlinear effects of school proximity on house prices in their area of interest.

While these studies made valuable contributions to understanding the impact of school proximity on house prices, a notable gap remains in the literature regarding the relationship between house prices and public school proximity in unique urban contexts, specifically the distance to public schools within the same school district where households reside. Hence, this study aims to address this gap by examining the relationship between house prices and distance from the nearest public school within the same school district in Allegheny County, Pennsylvania, thereby contributing to the existing literature on this topic. To achieve the aim of this study, the following research question is formulated:

"How does the proximity to the nearest public school within the same school district affect house prices in Allegheny County, Pennsylvania?"

Additionally, this study will expand on spatial heterogeneity, as initially explored by Sah et al. (2015), by investigating how the relationship varies across urban, suburban, and semi-rural area types within Allegheny County. This approach delves into relatively uncharted academic territory, providing new insights into how different area types might influence the relationship.

From a societal perspective, this study seeks to provide practical insights that can inform decision-making. Specifically, exploring the relationship between house prices and public school proximity can help stakeholders better understand the factors influencing house prices. Allegheny County is particularly relevant to study due to its unique morphology and its diverse composition, encompassing a mix of urban, suburban, and semi-rural areas. Moreover, Allegheny County's position as a major metropolitan area makes it a relevant case for understanding the relationship, providing insights for similar urban regions nationwide. Overall, the findings of this study provide insights to stakeholders about the potential implications of public school proximity on house prices, guiding informed decision-making.

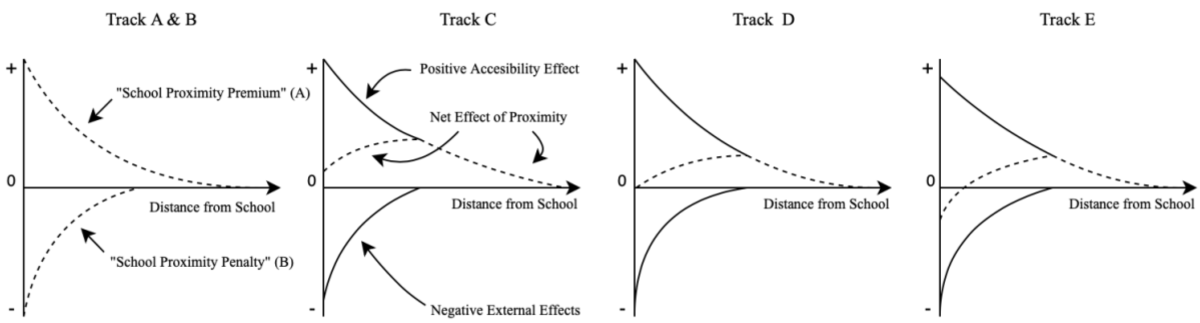
The rest of the paper is structured as follows: Section 2 reviews existing academic literature and provides a conceptual framework. Section 3 presents the empirical strategy, including the hedonic model and the Chow test. Section 4 presents and discusses estimation results. Finally, Section 5 concludes.

2. THEORY & LITERATURE REVIEW

While much of the existing literature has primarily focused on the influence of school quality on house prices, considerably less scholarly attention has been devoted to examining the impact of proximity to schools on house prices. To the extent that schools are often perceived as neighborhood amenities, with greater proximity offering easier access (i.e. less travel time and lower transportation costs). This convenience is attractive especially to households with school-age children as it enhances their daily routine and quality of life. Additionally living near schools may offer households a sense of security, knowing that their children are nearby and easily accessible (Sah et al., 2015; Huang & Dall'Erba, 2021). These factors collectively contribute to the perceived value of homes located in greater proximity to schools, thus potentially conferring a positive effect on house prices, as households may demonstrate a higher willingness to pay. Simultaneously, greater proximity to schools can impose negative effects on house prices due to increased noise levels, traffic congestion, petty crime, and vandalism. These negative externalities may diminish the desirability of living in greater proximity to schools, thus potentially conferring a negative effect on house prices (Sah et al., 2015; Huang & Dall'Erba, 2021). These competing effects contribute to the complexity of the relationship, where households weigh the perceived benefits against the perceived drawbacks and its net effect may vary depending on the specific context and the preferences of households.

In light of these conflicting effects, understanding the complex relationship between house prices and school proximity requires a comprehensive conceptual model that accounts for both competing effects associated with proximity to schools. This study presents a stylized graphical conceptualization (figure 1) based on the graphical conceptualization of Sah et al. (2015, Fig 2) and Li & Brown's figure for non-residential activity (Li & Brown, 1980, Fig 1). In the stylized graphical conceptualization, five possible scenarios "tracks" are outlined based on assumptions drawn from Li & Brown (1980). Initially, it is assumed that the positive price effect linked to accessibility decreases with distance from schools, while the negative price effect associated with negative external effects decreases with distance as well. Given that the negative effect from externalities decreases more rapidly with distance, it its expected that these negative effects will disappear much more rapidly as distance to school increases. This suggests that the positive accessibility effect curve exhibits a flatter trajectory compared to the negative external effects curve. In general, the net effect of proximity is the summation of the upper and lower curve and is visually represented by the dotted line in each track. An exception to this are tracks A & B, which are alternative tracks, where the curves do not represent the summation of upper and lower curves, but rather a clear net effect.

Fig 1 Stylized Graphical Conceptualization: Net Effect of Proximity



Track A illustrates a clear "school proximity premium" as the net effect remains positive, starting particularly positive for housing in immediate proximity to the school, then gradually declining towards zero as the distance from the school increases. Conversely, *Track B* illustrates a clear "school proximity penalty" as the net effect remains negative, starting particularly negative for housing in immediate proximity to the school, then gradually declining towards zero as the distance from the school increases. In *Track C*, the initial net effect is positive, as the positive accessibility effect outweighs the negative external effects for housing in close proximity, followed by a positive range of net proximity effects that gradually diminishes towards zero. In *Track D*, the net proximity effect starts at zero, where the negative external effects are counteracted by the positive accessibility effect for housing in close proximity, which is followed by a positive range of net proximity effects that gradually approach zero. In *Track E*, the initial net effect is negative, as the negative external effects outweigh the positive accessibility effect for housing in close proximity, followed by a positive range of net proximity effects that gradually diminishes towards zero. Assuming that the negative effect from externalities decreases more rapidly with distance compared to the positive price effect linked to accessibility, the peak net effect is observed at a moderate distance from public schools, as can be seen in Tracks C, D, and E. These different scenarios underscore the complex relationship, forming the theoretical base for analyzing the outcomes.

In the existing literature, there has been limited research on the relationship between house prices and school proximity. The earliest research on this topic by Emerson (1972) and Hendon (1973), who examine house prices in southern Minneapolis and Dallas utilizing hedonic pricing methods, find that house prices rise with increasing distance from the nearest school, proposing a "school proximity penalty". However, the findings of Emerson (1972) are viewed as suggestive rather than conclusive, as the statistical significance of the observed relationship was not firmly established. The study also included several spatial control variables to account for other geographical influences on house prices. Similar insignificant results are found by Li & Brown (1980), which they attribute to the accessibility of schools due to most houses being within a reasonable distance and the availability of bus transportation. Following this research, Guntermann & Cowell (1983) examine the effect of proximity to seven primary schools on house prices in Lubbock, Texas, and find a positive net effect with a range of 50 to 400 meters and an optimal distance of 230 meters. Rosier et al. (2001) measure the effect of both the size and proximity of primary schools on house prices using sales data of 4,300 single-family homes in Quebec, Canada from 1990 to 1991. They find that, on average, for each additional kilometer of distance from the school the price of a house decreases by $2,151. The authors determine that the ideal distance from the school, which maximizes the positive accessibility effect, falls within the range of 300 to 500 meters. Their analysis only includes one spatial control variable, namely park, which they discover has a positive amenity effect on house prices (see, amongst others, Crompton, 2001; and Crompton & Nicholls, 2019 for reviews on studies analyzing the impact of parks on house prices). Chin and Foong (2006) analyze how the accessibility of prestigious schools in Singapore affects house prices. They conclude that while school accessibility does add value to house prices, its impact is not as significant as other factors such as the reputation of the neighborhood and the tenure of the property. Owusu-Edusei et al. (2007) conduct a comprehensive study in Greenville, South Carolina, utilizing data from 3,732 single-family homes between 1994 and 2000. Their study stands out for controlling not only for distance to other amenities but also for school type and quality, providing a more comprehensive analysis of the relationship between school (proximity) and house prices. The findings reveal a net positive effect of school proximity on house prices. More specifically, the results suggest that homes located within 240 meters of an elementary school are priced approximately 8% to 13% higher compared to those situated between 240 meters and 3200 meters away from the school. Similarly, for middle schools, the value is approximately 12% higher for homes within the same distance ranges. On the contrary, high schools exhibit a negative effect on nearby house prices consistent with the notion of a "school proximity penalty", which the authors attribute to increased nighttime activity and lighting. Metz (2015) analyze home transactions in the Denver Public School District to explore the relationship between distance to assigned public schooling and house prices, considering elementary, middle, and high schools. Controlling for school quality and neighborhood characteristics, the study finds that house prices generally decrease with increased distance to schools. A 300-meter increase in distance leads to a 0.2 to 0.7 percent drop in house prices, with the weakest effect observed for middle schools. Homes very close to schools (less than 150 meters) are valued less than those 300 meters away, indicating a congestion effect. Additionally, homes just inside the walking cutoff distance for elementary schools (400 meters) are priced lower than those just outside it, and distances beyond 3.2 kilometers has no impact on prices. Sah et al. (2015) investigate the impact of school proximity on nearby residential housing, incorporating spatial heterogeneity into their analysis. They utilize a dataset of over 20,000 residential housing sales from 2010 and 2011 in San Diego County. In their study, they control for school quality and distances to various amenities such as freeways, downtown areas, the coast, libraries, malls, open spaces, and retail centers. Their findings reveal interesting patterns: in inland areas of San Diego County, there is a positive effect with proximity to public elementary schools but a negative effect for private elementary schools. However, in coastal areas, proximity to both types of schools exhibites negative effects on house prices. Despite these observed results, the authors do not pinpoint the specific sources of spatial heterogeneity. However, an intuitive explanation may lie in demographic and socio-economic disparities between San Diego County's inland and coastal areas. Inland regions, catering to families seeking affordability, may exhibit positive impacts on house prices near public elementary schools. Conversely, coastal areas, appealing to retirees or luxury property seekers, may show negative effects near both public and private elementary schools due to differing priorities and amenities. Huang and Hess (2018) employe quantile regression for their research in Oshkosh, Wisconsin, and find that on average the median sales price decreases as the distance to the nearest elementary, middle, and high schools increases. Finally, Huang & Dall'Erba (2021) examine the relationship between proximity to secondary schools and house prices within four school enrollment zones in Auckland, New Zealand, utilizing both hedonic pricing and quantile regression methods. Their results suggest that in highly desirable school districts, house prices tend to increase as the distance to the school decreases, but this trend reverses beyond 4 kilometers. Notably, this effect is most pronounced for houses at the lower end of the price distribution. Conversely, in other school districts, proximity to the school is associated with lower house prices. Although the authors do not explain the pronounced effect for houses at the lower end of the price distribution, one obvious reason could be that families with lower budgets prioritize factors like school proximity due to cost, and time savings.

Overall, the discussed literature highlights a varied relationship between school proximity and house prices, with factors such as school type, quality, neighborhood characteristics, and spatial heterogeneity playing significant roles in influencing house prices.

3. METHODOLOGY & DATA

3.1 Methodology

Through the application of the hedonic pricing model, this study aims to investigate the potential impact of school proximity on house prices. The hedonic pricing model serves as an analytical framework, examining the intrinsic value of a property's characteristics, encompassing structural, neighborhood, spatial, and environmental characteristics, through the analyses of housing transaction data. Rosen (1974) introduced the idea of decomposing market goods like housing into nonmarket components, enabling the determination of a price schedule for these attributes. Unlike traditional market dynamics, where sellers pursue profit maximization and buyers seek utility maximization, the hedonic pricing method operates within an implicit framework. This framework infers prices from consumption and production decisions, with housing costs being seen as a function of its structural, neighborhood, and environmental characteristics (Cheshire & Sheppard, 1998). The hedonic price function is written as follows:

$$(1)\ Ph = fh(s, n, e)$$

The price of a property (Ph) depends on various housing characteristics (fh), which encompasses traditional hedonic features like structural components (s), neighborhood characteristics (n), and environmental characteristics (e). Through regression analysis of housing transaction sales prices and the characteristics of sold properties, this approach reveals the marginal valuations or "implicit" prices associated with each housing component. Hedonic theory suggests that, holding all other factors constant, properties in closer proximity to schools are expected to command higher prices than those located further away.

This study utilizes two different approaches to model the impact of public school proximity on house prices: one with a single coefficient for distance and another with distance category dummies. By employing a single coefficient for distance, this study aims to determine both the direction and magnitude of the relationship. However, to achieve a more comprehensive assessment of the relationship, an alternative model specification is introduced. This model incorporates measures of proximity to public schools divided into discrete distance intervals. This approach is motivated by the need to address the (partially) nonlinear nature of the variable, as indicated by issues with the functional form in the continuous model. This method aligns with previous research on the impact of proximity to schools on house prices and provides a solution for the nonlinearity of the variable. Furthermore, using discrete distance intervals instead of a quadratic distance variable allows the model to capture potential non-linearities more flexibly and accurately. These intervals provide a better representation of the relationship between proximity to schools and house prices, accommodating potential nonlinear patterns that a quadratic distance variable may overlook. Additionally, this approach is more interpretable and aligns with the intuitive understanding that the impact of proximity to schools on house prices may vary across different distance ranges. The distance categories, originally derived from a study by Sah et al. (2015), underwent a slight adjustment in this study due to the use of centroids instead of outlines of public schools. The distance categories are delineated as follows: 0-200 meters, followed by 201-500 meters, then 501-800 meters, etc., with each subsequent category representing an increment of 300 meters up to 1700 meters. Beyond the 1700-meter mark, housing locations serve as the reference category for the distance category dummy. The empirical framework incorporates variables aggregated at the school district level. This methodological decision is grounded in the assumption that amenities within school districts share similarities, allowing for variations in prices to primarily reflect property-specific characteristics and other important variables, including distance to the nearest public school within the same school district where the property is situated. Furthermore, aggregating variables at the school district level partially addresses the issue of not being able to directly control for school quality. Including the school district variable essentially captures the overall characteristics and amenities associated with that specific school district, which indirectly reflects differences in school quality. One could argue that this approach may not fully account for variations in school quality across different schools within the same district. However, previous research found that households are generally willing to pay a premium to reside in districts with higher-quality schools, ceteris paribus. As a result, it is assumed that not accounting for variations in school quality across different schools within the same district will not pose a significant problem. Acknowledging that price fluctuations may stem from economic changes over time, time-related fixed effects are incorporated to address temporal dynamics. Consequently, the timing of transactions is included in the model to capture the temporal intricacies inherent in the housing market. The hedonic multivariable regression model utilized in this study can be represented as follows:

$$lnP_{ijt} = \alpha + \beta_1 \text{DistancePublicSchool}_{ijt} + \beta_2 \text{X}_{ijt} + \beta_3 \text{S}_{ijt} + \gamma_1 \text{L}_j + \gamma_2 \text{L}_t + \epsilon_{ijt}$$

The dependent variable, lnP_{ijt} represents the natural logarithm of the price of property *i* within school district j in sale year *t*. The main independent variable, $\text{DistancePublicSchool}_{ijt}$ denotes the distance from a given observation to the nearest public school within school district j, measured in kilometers and calculated based on the public school's centroid. X_{ijt} incorporates property attributes of property *i* within school district *j* sold in year *t*, including property age, property size, number of bedrooms, property condition, property type, if the property is an apartment or not, if the property has a basement or not, and if the property has air conditioning or not[1]. S_{ijt} incorporates a spatial control variable of property *i* within school district j sold in year t, included to enhance the robustness of the findings. This variable encompasses distance to the nearest Allegheny County state park, measured in kilometers from each observation to the nearest point along the border of the nearest state park. L_j encompasses neigborhood fixed effects, consisting of dummy variables for school districts, while L_t represents year-specific fixed effects to account for temporal variations, corresponding to the transaction year of property *i*. Lastly, ϵ_{ijt} represents the error term. Both the dependent variable lnP and the independent variable "finishedlivingarea" were subject to natural logarithm transformations because of their initial skewness and a large spread of data (see Appendix B and C). As highlighted by Brooks & Tsolacos (2010), this adjustment ensures that

[1] There is no multicollinearity issue in the model, as indicated by the VIF value being below 5.00; Appendix B provides the correlation matrix for additional information.

the model's relationships are more linear, bringing the distribution closer to a normal distribution and facilitating achieving linearity between the variables (see Appendix D and E).

Following the application of the hedonic pricing model to analyze the impact of school proximity on house prices, this study proceeds to investigate spatial heterogeneity using the Chow test. The Chow test is integral to the analysis, examining spatial heterogeneity in the relationship (Chow, 1960). It detects structural changes by comparing regressions conducted jointly with those conducted separately, focusing on identifying shifts in coefficients across different area types. Hypotheses are formulated, with the null hypothesis proposing constant coefficients across the dataset, and the alternative suggesting variation. Pooled regression analysis covers the overall dataset and specific area types, followed by the computation of the F-value for the Chow test. This study aims to evaluate whether the coefficient effects of the relationship differ across urban, suburban, and semi-rural areas, indicating spatial heterogeneity. School districts have been categorized into urban, suburban, or semi-rural area types based on their respective characteristics, including population density, infrastructure, and land use patterns. The full categorization can be found in Appendix A.

3.3 Data

The datasets used in this study are obtained from the Western Pennsylvania Regional Data Center (WPRDC), an open-access database overseen by the University of Pittsburgh Center for Urban and Social Research, and funded jointly by Allegheny County and the City of Pittsburgh, (WPRDC, 2020). Four datasets were utilized in this study: the first one containing 575,736 digitized property assessment records with parcel centroids, the second containing the centroid locations of 276 Allegheny County Public Schools, the third delineating the outlines of the 45 Allegheny County school district boundaries, and a fourth delineating the outlines of the 10 Allegheny County State Parks.

The dataset, which includes digitized property assessment data with parcel centroids, along with the dataset containing centroid locations of 276 Allegheny County Public Schools and the dataset delineating the outlines of the 45 Allegheny County school district boundaries, enables geocoding at the property level by integrating geographic coordinates through Geographic Information System (GIS). This allows for the calculation of distances, measured in meters, from a specific observation to the nearest public school within the respective school district where the observation is located, serving as our main independent variable. The same spatial technique is applied to the dataset delineating the outlines of the 10 Allegheny County state parks to calculate the distance from a specific observation to the nearest point along the border of the nearest state park, which is utilized as a spatial control variable in the analyses.

The data cleaning process involved several steps to ensure the quality and integrity of the dataset. Initially, non-residential observations were removed, followed by the exclusion of residential properties with deviating uses. Observations, where the sales price was not representative of the current market value during the sale, were also filtered out. Further cleaning involved removing observations with missing or false values for important variables. A new variable, "saleyear," was generated to capture the year of the sale, and observations prior to the year 2010 were removed. It is important to note that the dataset containing information on public schools did not include specific details on the year each school was founded. While historical information was consulted, the latest record for a school established was found to be in 1998. However, it should be acknowledged that there may be schools that were established later, but this information was not readily available. As a conservative approach, observations prior to the year 2010 were selected as a cut-off point. This ensures that most if not all, public schools in the dataset were likely founded prior to the year 2010. Furthermore, prior to transforming the dependent variable lnP and the independent variable "finishedlivingarea", I took precautions to ensure the robustness of the statistical analysis by omitting the first and last 1% of transactions from the dataset to mitigate the potential influence of outliers on the results. After completing the data-cleaning process, I was left with 38,266 observations for analysis.

3.4 Descriptive Statistics

TABLE 1. DESCRIPTIVE STATISTICS

Variable	Mean	St. Dev	Min	Max
Dependent Variable				
Sales Price *(in $)*	197,276	125,908	23,900	849,000
LN Sales Price	12.00	0.63	10.08	13.65
Independent Variable				
Distance to Nearest Public School *(in km)*	1.258	1.000	0.032	8.376
Distance to Nearest Public School 0-200 m	0.03	0.17	0	1
Distance to Nearest Public School 201-500 m	0.17	0.38	0	1
Distance to Nearest Public School 501-800 m	0.20	0.40	0	1
Distance to Nearest Public School 801-1100 m	0.16	0.37	0	1
Distance to Nearest Public School 1101-1400 m	0.12	0.32	0	1
Distance to Nearest Public School 1401-1700 m	0.09	0.28	0	1
Distance to Nearest Public School > 1701 m	0.24	0.42	0	1
Control Variables				
Finished Living Area *(in square feet)*	1,609.3	602.2	720	3,949
LN Finished Living Area	7.32	0.35	6.58	8.28
Property Age *(in years)*	71.73	27.14	3	269
Number of Bedrooms	2.99	0.80	0	9
Airconditioning (1) or not (0)	0.70	0.46	0	1
Basement (1) or not (0)	0.94	0.24	0	1
Apartment (1) or not (0)	0.06	0.24	0	1
Physical Condition ‘Unsound’ *(1=yes)*	0.00	0.01	0	1
Physical Condition ‘Very Poor’ *(1=yes)*	0.00	0.02	0	1
Physical Condition ‘Poor’ *(1=yes)*	0.00	0.07	0	1
Physical Condition ‘Fair’ *(1=yes)*	0.08	0.27	0	1
Physical Condition ‘Average’ *(1=yes)*	0.77	0.42	0	1
Physical Condition ‘Good’ *(1=yes)*	0.13	0.33	0	1
Physical Condition ‘Very Good’ *(1=yes)*	0.01	0.12	0	1
Physical Condition ‘Excellent’ *(1=yes)*	0.00	0.06	0	1
Distance to Nearest State Park *(in km)*	6.156	3.225	0.010	15.383
Total Number of Observations = 38,266				

Table 1 provides a comprehensive overview of the descriptive statistics for the dataset, encompassing 38,266 observations. The dependent variable, sales price, averages $197,276 with a substantial standard deviation of $125,908, highlighting significant variability in house prices. When log-transformed, the mean is 12.00 with a standard deviation of 0.63, suggesting a more normalized distribution of prices. The key independent variable, distance to the nearest public school, averages 1.258 km with a standard deviation of 1.000 km, ranging from as close as 0.032 km to as far as 8.376 km. This variable is further broken down into categorical distances, providing a better understanding of proximity to public schools. Notably, a small percentage (3%) of homes are within 200 meters of a public school, while 17% fall between 201 and 500 meters, 20% between 501 and 800 meters, 16% between 801 and 1100 meters, 12% between 1101 and 1400 meters, 9% between 1401 and 1700 meters, and the largest group (24%) is located more than 1700 meters away. The control variables offer additional context about the properties. The average finished living area is 1,609.3 square feet, with a log-transformed mean of 7.32. Properties have an average age of 71.73 years, reflecting a mix of historical and newer homes. Most homes have around three bedrooms, and 70% have air conditioning. Basements are common, present in 94% of homes, while 6% are categorized as apartments. Physical condition categories reveal that the majority of homes are in 'Average' condition (77%), with smaller proportions in 'Good' (13%), ‘Fair’ (8%), and ‘Very Good’ (1%). A negligible percentage of homes are deemed to be in 'Unsound', 'Very Poor', ‘Poor’, and ‘Excellent’ condition. Lastly, the distance to the nearest state park averages 6.156 km, with a considerable range from 0.010 km to 15.383 km, indicating a varied proximity to state parks.

Fig 2. Spatial Distribution of Observations, Public Schools, and School Districts

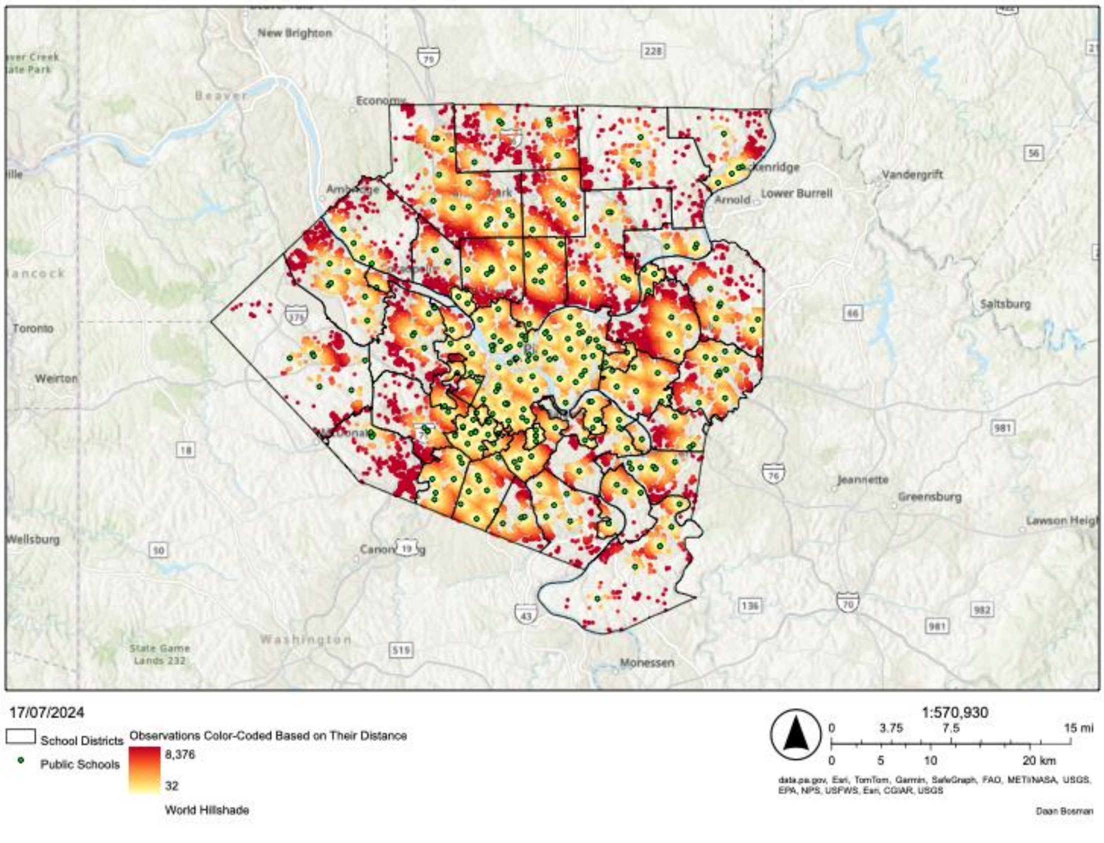


Figure 2 illustrates the spatial distribution of observations, public schools, and school districts within Allegheny County. Observations are color-coded based on their distance to the nearest public school, with dark red indicating higher values (greater distances) and light yellow indicating lower values (shorter distances). School districts are delineated by black polygons, and public schools are marked by bright green dots. A higher density of observations is observed closer to the center of Pittsburgh. Conversely, the density of observations is lower and more dispersed as distance from Pittsburgh increases. A similar pattern is evident concerning the number of public schools, which have a higher concentration closer to the center. Notably, homes in the urban center are generally closer to public schools, as indicated by the light yellow coloration, compared to surrounding areas, reflecting the need for accessible education in densely populated regions. See Appendix H for a zoomed-in image of figure 2.

4. RESULTS & DISCUSSION

4.1 Main Results

Three OLS regression models are employed to examine the relationship between house prices and distance to the nearest public school within the same school district in Allegheny County. These models highlight the association between house prices and proximity to public schools, without implying causation. Models 1 and 2 are used to determine both the direction and magnitude of the relationship, while model 3 provides additional insights by accounting for the nonlinearity in the relationship. Table 2 provides model estimates for the three OLS models. By using natural logarithms, changes can be interpreted as percentages rather than absolute values. To address the issue of heteroskedasticity and thereby ensure reliable statistical conclusions, robust standard errors were used consistently throughout the regression analysis.

Model 1 examines how the natural logarithm of house prices varies with the distance to the nearest public school within the same school district, measured in kilometers. The model does not control for property characteristics or neighborhood and year-fixed effects. The model's R-squared of 0.000 suggests that the variance in house prices is not explained by the distance to the nearest public school. Additionally, the insignificant positive relationship, with a coefficient of 0.002 and a p-value of 0.456, would naively imply that, on average, for each kilometer of increased distance from the nearest public school within the same school district, house prices in Allegheny County increase by around 0.2%, ceteris paribus. However, these results are suggestive rather than conclusive as the observed relationship lacks strong statistical significance, indicating that the model is naive due to its omission of relevant control variables.

Model 2 refines the analysis by incorporating controls for property characteristics, neighborhood, and year-fixed effects. This enhanced model provides a clearer understanding of the association between house prices and school proximity. With an R-squared value of 0.713, the model demonstrates that 71.3% of the variance in house prices can be explained by the included variables. The significant negative relationship observed in the model, with a coefficient of -0.029 and a p-value of 0.000, suggests that, on average, for each kilometer of increased distance from the nearest public school within the same school district, house prices in Allegheny County decrease by 2.9%, ceteris paribus. This result contrasts starkly with the insignificant positive relationship identified in Model 1, highlighting the importance of controlling for relevant factors in accurately understanding the true nature of the relationship between house prices and proximity to public schools in this context.

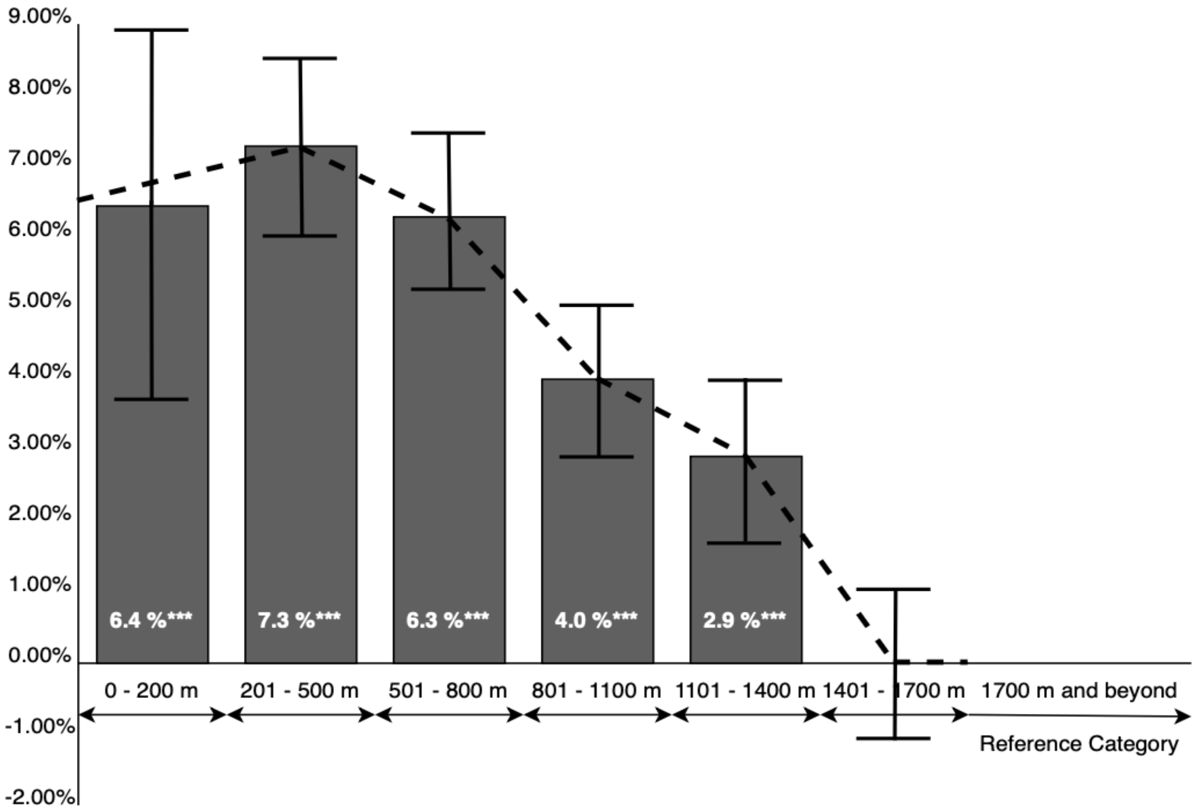
Model 3 incorporates measures of proximity to public schools divided into discrete distance intervals. This approach addresses the partially nonlinear nature of the relationship between house prices and school proximity, as suggested by issues with the functional form in the continuous models. By using discrete intervals, Model 3 can more accurately capture the varying impact of proximity across different distances. The R-squared value remains unchanged from Model 2 at 0.713, indicating that the model explains 71.3% of the variance in house prices. The estimation results from Model 3 reveal significant effects of proximity to public schools on house prices when compared to properties more than 1700 meters away. Specifically, properties within 0-200 meters of a public school see a 6.4% increase in house prices (p-value: 0.000), those within 201-500 meters experience a 7.3% increase (p-value: 0.000), those within 501-800 meters have a 6.3% increase (p-value: 0.000), those within 801-1100 meters see a 4.0% increase (p-value: 0.000), and those within 1101-1400 meters have a 2.9% increase (p-value: 0.000). Conversely, the distance interval of 1401-1700 meters shows an insignificant effect on house prices, with a negligible coefficient and a p-value of 0.946. These results indicate that, in general, closer proximity to public schools is associated with higher house prices, with the effect diminishing and becoming insignificant beyond 1400 meters, ceteris paribus. Figure 3 graphically visualizes the estimation results from Model 3, with a dashed line illustrating the trend and error bars showing the 95% confidence intervals.

TABLE 2. OLS ESTIMATION RESULTS OF LN SALES PRICE

	Model 1	Model 2	Model 3
Distance to Nearest Public School in km (continuous)	0.002 (0.72)	-0.029*** (-13.20)	
Distance to Nearest Public School (categorical)			
Distance to Nearest Public School 0 - 200 m			0.064*** (4.76)
Distance to Nearest Public School 201 - 500 m			0.073*** (11.63)
Distance to Nearest Public School 501 - 800 m			0.063*** (11.25)
Distance to Nearest Public School 801 - 1100 m			0.040*** (7.13)
Distance to Nearest Public School 1101 - 1400 m			0.029*** (5.04)
Distance to Nearest Public School 1401 - 1700 m			4.277e-04 (0.07)
Control Variables	No	Yes	Yes
Neighborhood Fixed Effects	No	Yes	Yes
Year Fixed Effects	No	Yes	Yes
Number of Observations	38,266	38,266	38,266
R^2	0.000	0.713	0.713

T-Statistics in Parentheses
* = p < 0.1, ** = p < 0.05, *** = p < 0.01
For Model 3, Beyond the 1700-meter mark, housing locations serve as the reference category.
Note: Robust Standard Errors Used

Fig 3. Graphical Visualisation of Model 3 Results



4.2 Spatial Heterogeneity Across Area Types

The analysis extends beyond ordinary least squares (OLS) regression by employing the Chow test to examine potential variations in the relationship across urban, suburban, and semi-rural areas within the dataset. This investigation allows us to delve into spatial heterogeneity, exploring how the relationship might differ across distinct geographical contexts. Initially, the restricted model (R RSS) encompassed all area types, serving as a baseline for comparison. Subsequently, three separate regressions are conducted targeting specific area subgroups: urban (U RSS^1), suburban (U RSS^2), and semi-rural (U RSS^3), each serving as an unrestricted model. Considering that the number of school districts varies within each area type, including school district fixed effects in each model leads to a varying number of variables. This discrepancy poses challenges for the Chow test comparison, as it affects the degrees of freedom and could introduce bias into the results. Therefore, we excluded school districts as neighborhood fixed effects in our models. Consequently, the R-squared of the restricted model (0.606) is understandably lower than that of Model 2 (0.713). Table 3 presents the estimation results for the Chow test. Across most models, significant evidence is found of a negative relationship between the natural logarithm of house prices and the distance to the nearest public school within the same school district. However, the coefficient for the semi-rural subgroup (U RSS^3) was found to be insignificant. The significance of the Chow test was assessed using the F-statistic, derived from the residuals of the models. With 38,266 observations (n), 27 parameters (k), and 3 groups (g), the resulting F-statistic is 68.500. This value significantly exceeds the critical F-statistic range of 1.318 to 1.394 from the F-distribution table, providing strong evidence to reject the null hypothesis of the Chow test with the null hypothesis proposing constant coefficients across the dataset. Therefore, results indicate significant differences between the coefficients of the unrestricted models and the restricted model, thereby suggesting spatial heterogeneity across urban, suburban, and semi-rural areas.

The estimation results of the first unrestricted model suggest that, on average, for each kilometer of increased distance from the nearest public school within the same school district, house prices in urban areas in Allegheny County decrease by 14.0%, ceteris paribus. The estimation results of the second unrestricted model suggest that, on average, for each kilometer of increased distance from the nearest public school within the same school district, house prices in suburban areas in Allegheny County decrease by 1.2%, ceteris paribus. The estimation results of the third unrestricted model suggest that, on average, for each kilometer of increased distance from the nearest public school within the same school district, house prices in semi-rural areas in Allegheny County decrease by 0.08%, ceteris paribus. However, the coefficient for the semi-rural subgroup ($U\ RSS^3$) is suggestive rather than conclusive, as the observed relationship lacks strong statistical significance. The estimation results of the urban and suburban models indicate a 12.8 percentage point difference, suggesting that the effect of proximity to public schools on house prices is much more pronounced in urban areas compared to suburban areas.

TABLE 3. OLS ESTIMATION RESULTS OF CHOW-TEST OF LN SALES PRICE

	R RSS	$U\ RSS^1$ Urban	$U\ RSS^2$ SubUrban	$U\ RSS^3$ Semi-Rural
Distance to Nearest Public School in km's (continuous)	-0.023***	-0.140***	-0.012***	-0.008
	(-11.0)	(-8.67)	(-4.64)	(-0.30)
Control Variables	Yes	Yes	Yes	Yes
Neighbourhood Fixed Effects	No	No	No	No
Year Fixed Effects	Yes	Yes	Yes	Yes
Number of Observations	38,266	6,244	28,840	3,182
Residual Sum of Squares	5,956	1,360	3,687	383
Number of Independent Variables	27	27	27	27
Adjusted R^2	0.606	0.594	0.650	0.666

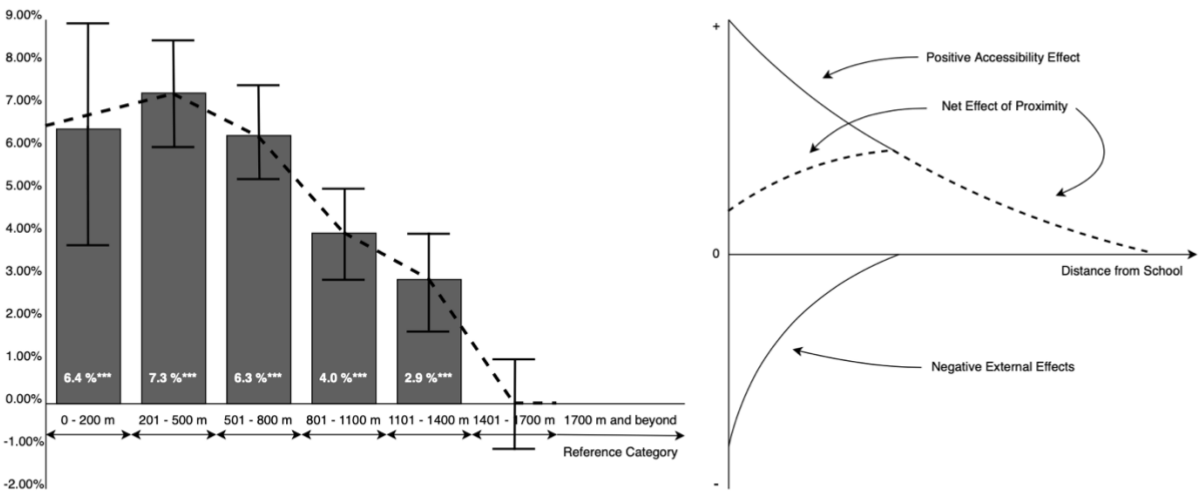
T-Statistics in Parentheses
F ~ (54, 38185)
* = p < 0.1, ** = p < 0.05, *** = p < 0.01

4.3 Discussion

4.3.1 Discussion of Main Results

The results of this analysis provide insights into the relationship between house prices and proximity to the nearest public school within the same school district in Allegheny County. The results from model 1 are inconclusive due to the lack of strong statistical significance and the absence of controls for property characteristics, neighborhood, and year-fixed effects. Consequently, these results will not be discussed. The significant negative relationship observed in Model 2 suggests that, on average, for each kilometer of increased distance from the nearest public school within the same school district, house prices in Allegheny County decrease by 2.9%, ceteris paribus. In other words, properties that are closer to public schools tend to have higher prices, indicating that proximity to schools is a valued neighborhood amenity among homebuyers in Allegheny County. This finding implies a "school proximity premium," where the convenience and accessibility benefits associated with living near a public school outweigh the potential negative externalities, such as increased noise levels and traffic congestion, thereby enhancing the attractiveness and value of these properties. The results from Model 3 provide empirical evidence supporting the theoretical framework outlined in section 2. Specifically, the findings align with *Track C* from the conceptual model, which illustrates a scenario where the initial net effect of proximity to schools is positive, with the peak net effect observed at a moderate distance from public schools, before gradually diminishing towards zero as distance from the schools increases. In Track C, the positive accessibility effect outweighs the negative external effects for housing in close proximity to schools, leading to increased house prices. However, as the negative effect from externalities decreases more rapidly with distance compared to the positive price effect linked to accessibility, the highest increase in house prices is observed between the 200 and 500-meter interval. This suggests that households in this range experience accessibility benefits with almost negligible negative effects. However, beyond the 500-meter mark, the magnitude of the accessibility effect starts to decrease, eventually becoming insignificant beyond the 1400-meter distance interval. This diminishing effect suggests that while proximity to public schools remains a desirable attribute, its impact on house prices diminishes as distance increases. This phenomenon is consistent with the idea that the positive benefits of proximity diminish and are offset by other factors as distance from the school increases. Notably, the 95% confidence intervals indicate a wider range of potential effects for each interval, suggesting that the true effect could result in an alignment with a different track. Overall, the alignment of model 3 results with *Track C* underscores the complex relationship between house prices and school proximity, highlighting the importance of considering both the positive accessibility effects and negative externalities associated with proximity to schools when analyzing housing dynamics. Figure 4 combines Figure 3 with Track C from Figure 1, enabling a direct comparison between the empirical results and the conceptual model.

Fig 4. Graphical Visualisation of Model 3 Results and Track C from Stylized Conceptual Model



4.3.2 Discussion of Spatial Heterogeneity Across Area Types

The Chow test results highlight significant differences between the coefficients of the unrestricted models and the restricted model, indicating spatial heterogeneity across urban, suburban, and semi-rural areas. Consistently, each model suggests a negative relationship, aligning with the results from model 2. However, there are notable variations in the relationship across different area types in Allegheny County. Urban areas exhibit a substantial decrease of 14.0% for each kilometer of increased distance from the nearest public school within the same school district, contrasting with more modest decreases of 1.2% in suburban areas and 0.08% in semi-rural areas. Nonetheless, the observed relationship in semi-rural areas lacks strong statistical significance indicating limited impact on house prices in these areas.

The substantial decrease in house prices observed in urban areas might be attributed to factors such as higher population density, limited space, and increased competition for housing closer to public schools. Additionally, in urban areas, negative externalities like noise and traffic congestion are often perceived as inherent aspects of the environment rather than specific to certain locations. Residents may become accustomed to these urban challenges and may not attribute them solely to proximity to schools. As a result, the positive accessibility benefits of living near schools may stand out more prominently. Furthermore, public schools in urban areas often serve as hubs for various community amenities beyond education alone. They may include facilities such as sports fields, playgrounds, community centers, and urban green spaces. The presence of these additional amenities near public schools further reinforces the importance of proximity in influencing house prices, especially in densely populated areas.

The more modest decreases in house prices in suburban areas may reflect a different set of preferences and priorities among homebuyers. Suburban neighborhoods typically offer more space, quieter surroundings, and a perceived sense of safety, which could mitigate the negative impact of distance from public schools on house prices. Additionally, suburban areas often have better-developed transportation infrastructure, making it easier for residents to commute to schools or other amenities. As households in suburban areas are accustomed to covering larger distances it diminishes the relative impact of proximity to schools on house prices. Furthermore, since homebuyers in suburban areas prioritize quieter surroundings, negative externalities such as noise and traffic congestion may be more heavily weighted, leading to a more modest overall impact of public school proximity on house prices.

The results suggesting a minimal decrease in house prices in semi-rural areas lacks strong statistical significance. This insignificance implies that factors such as privacy, natural landscapes, and a slower pace of life in semi-rural areas may diminish the importance of proximity to public schools for homebuyers. Additionally, the dispersed population in semi-rural areas results in less competition for housing, reducing the pressure on prices to decrease with distance from schools. Moreover, households in semi-rural areas are even more accustomed to covering larger distances compared to those in suburban areas, further diminishing the relative impact of proximity to schools on house prices. The lack of statistical significance suggests that other factors not captured in the analysis may be influencing house prices in these areas. Overall, the analysis reveals that proximity to public schools has a varying impact on house prices depending on the area type within Allegheny County, highlighting the importance of local context and varying homebuyer preferences in shaping these dynamics.

4.3.3 Limitations

While this study makes a significant contribution, it's crucial to recognize its limitations. One limitation of this study is its focus on examining the relationship between house prices and proximity to the nearest public school within the same school district in a distinct urban setting. Firstly, analyzing within the same school district restricts the generalizability of the findings, particularly considering that the United States is one of the few countries where public school choice is generally confined to the school districts where households reside. Furthermore, when combined with the unique urban context of Allegheny County, the generalizability of our results is further limited to, at most, similar urban regions within the US. Although the vast majority of students (90%) attend public schools (National Center for Education Statistics, n.d.), meaning private schools cater to a small proportion of the student population, not controlling for the presence of private schools could bias the estimates, potentially overstating the effect of public school proximity. Additionally, employing discrete distance intervals to examine the relationship captures the nonlinear dynamics but faces challenges due to the loss of precise distance information. Discrete distance intervals may obscure subtle gradients in the relationship, while arbitrary interval choices can introduce uncertainty and bias as different cutoff points can yield different results. Moreover, although the utilization of neighborhood fixed effects provides a coarser level of granularity compared to more detailed fixed effects such as those at the zip-code level, it still absorbs an amount of spatial variance, thereby limiting the explanatory power of distance intervals. Another limitation lies in the utilization of centroids rather than outlines, a concern already mentioned by Rosiers et al. (2001). Although the distance intervals in this study underwent a slight adjustment to account for the use of centroids, it can still introduce a measurement issue as it fails to entirely adjust for public school size. For instance, a distance of 150 meters from the centroid of a small public school might approximate 100 meters from the public school's perimeter. As a result, the accuracy of proximity measurement to public schools may vary depending on the size of the school property, leading to potential bias and inaccuracies in the results. Furthermore, despite adopting a conservative cutoff approach in 2010, it remains possible that public schools were established after this date. This potential occurrence could introduce minor temporal discrepancies, thus possibly biasing the results. The omission of other relevant variables such as neighborhood amenities and the specific timeframe may further limit the scope and applicability of this study's findings. Lastly, given the cross-sectional approach of this study, the findings aim to identify associations rather than establish causal relationships. For causal inference, it is important to highlight that existing literature provides more robust methodologies, which have been applied in this field starting from studies like Black (1999) to more recent studies like the one conducted by Peng et al. (2021).

4.3.4 Implications and Future Research

This study holds noteworthy implications across societal, professional, and policy domains. It empowers households to make informed housing decisions based on educational access. Real estate professionals can utilize these findings to better serve their clients, offering tailored advice that aligns with educational priorities. Urban planners and policymakers can also benefit from this study, shaping strategies to address housing and educational needs effectively. Since households value proximity to public schools, strategic placement is crucial. Public schools should be conveniently located to improve access to education. Consequently, thoughtful public school placement can improve neighborhood quality and ensure equitable access to education, benefiting both communities and property values. Overall, the findings of this study provide insights into the potential implications of public school proximity on house prices, guiding more informed and strategic decision-making processes.

Future research could focus on several key areas and countries to enhance our understanding of the relationship between house prices and school proximity. Delving deeper into spatial heterogeneity could shed light on relatively unexplored territory, providing insights into how variations in geographical factors influence the relationship between house prices and school proximity. Additionally, incorporating measures of school quality into future research could elucidate whether differing levels of school quality, in combination with proximity, yield varying results. Moreover, considering the presence of private schools and distinguishing between elementary and secondary public schools in future studies could further enrich our understanding of the relationship. Furthermore, qualitative research methods such as interviews and focus groups could complement quantitative analyses by exploring subjective perceptions and experiences related to school proximity and housing decisions.

5. CONCLUSION

This study employed a comprehensive methodological approach, combining OLS regression models and the Chow test, to investigate the relationship between house prices and distance from the nearest public school within the same school district in Allegheny County, Pennsylvania. Through robust analysis, this study has established the direction and magnitude of this relationship, revealing a significant negative association between house prices and distance from public schools, with homes closer to public schools commanding higher prices. More specifically, the findings suggest that, on average, each kilometer further from the nearest public school within the same school district corresponds to a 2.9% decrease in house prices in Allegheny County, ceteris paribus. Moreover, the findings indicate that the impact of proximity varies across different distances, where the initial net effect of proximity to schools is positive, with the peak net effect observed at a moderate distance, before gradually diminishing towards zero as distance from the schools increases. Importantly, the findings align closely to track C from the conceptualized model suggesting that the model accurately captures these effects and provides a more precise understanding of how school proximity influences house prices across varying distances. Furthermore, this study has identified spatial heterogeneity in the relationship across urban, suburban, and semi-rural areas, highlighting the complex dynamics shaping housing markets in diverse geographical contexts. Urban areas exhibit a substantial decrease of 14.0% for each kilometer of increased distance from the nearest public school within the same school district, contrasting with more modest decreases of 1.2% in suburban areas and 0.08% in semi-rural areas. However, the observed relationship in semi-rural areas lacks strong statistical significance. Overall, this study contributes to existing literature by providing new insights into the relationship between house prices and public-school proximity, particularly regarding spatial heterogeneity in the relationship.

REFERENCES

Black, S. E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal Of Economics*, *114*(2), 577–599. https://doi.org/10.1162/003355399556070

Brasington, D., & Haurin, D. R. (2006). Educational Outcomes and House Values: A Test of the value added Approach*. *Journal Of Regional Science*, *46*(2), 245–268. https://doi.org/10.1111/j.0022-4146.2006.00440.x

Brooks, C., & Tsolacos, S. (2010). *Real estate modelling and forecasting*. https://doi.org/10.1017/cbo9780511814235

Cheshire, P., & Sheppard, S. (1998). Estimating the demand for housing, land, and neighbourhood characteristics. *Oxford Bulletin of Economics and Statistics*, *60*(3), 357–382. https://doi.org/10.1111/1468-0084.00104

Chin, H. C., & Foong, K. W. (2006). Influence of school accessibility on housing values. *Journal of Urban Planning and Development*, *132*(3), 120–129. https://doi.org/10.1061/(asce)0733-9488(2006)132:3(120

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*(3), 591. https://doi.org/10.2307/1910133

Downes, T. A., & Zabel, J. E. (2002). The impact of school characteristics on house prices: Chicago 1987–1991. *Journal Of Urban Economics*, *52*(1), 1–25. https://doi.org/10.1016/s0094-1190(02)00010-4

Crompton, J. L. (2001). The Impact of parks on Property Values: A review of the Empirical evidence. *Journal of Leisure Research*, *33*(1), 1–31. https://doi.org/10.1080/00222216.2001.11949928

Crompton, J. L., & Nicholls, S. (2019). Impact on property values of distance to parks and open spaces: An update of U.S. studies in the new millennium. *Journal of Leisure Research*, *51*(2), 127–146. https://doi.org/10.1080/00222216.2019.1637704

Emerson, F. C. (1972). Valuation of residential amenities: an econometric approach. *The Appraisal Journal*, *40*, 268–278.

Figlio, D. N., & Lucas, M. E. (2004). What's in a Grade? School Report Cards and the Housing Market. *The American Economic Review*, *94*(3), 591–604. https://doi.org/10.1257/0002828041464489

Gibbons, S., & Machin, S. (2008). Valuing school quality, better transport, and lower crime: evidence from house prices. *Oxford Review Of Economic Policy*, *24*(1), 99–119. https://doi.org/10.1093/oxrep/grn008

Guntermann, K. L., & Colwell, P. F. (1983). Property values and accessibility to primary schools. *Real Estate Appraiser and Analyst*, *49*, 62-68.

Hendon, W. S. (1973). Property values, schools, and Park-School combinations. *Land Economics*, *49*(2), 216. https://doi.org/10.2307/3145285

Huang, P., & Hess, T. (2018). Impact of Distance to School on Housing Price: Evidence from a Quantile Regression. *The Empirical Economics Letters*, *17*(2). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3096145

Huang, Y., & Dall'Erba, S. (2021). Does proximity to school still matter once access to your preferred school zone has already been secured? *The Journal of Real Estate Finance and Economics*, *62*, 548–577.

Li, M. M., & Brown, H. J. (1980). Micro-Neighborhood Externalities and Hedonic Housing Prices. *Land Economics*, *56*(2), 125. https://doi.org/10.2307/3145857

Machin, S. (2011). Houses and schools: Valuation of school quality through the housing market. *Labour Economics*, *18*(6), 723–729. https://doi.org/10.1016/j.labeco.2011.05.005

Metz, N. E. (2015). Effect of Distance to Schooling on Home Prices. ˜ *The œ Review Of Regional Studies/The Review Of Regional Studies*, *45*(2). https://doi.org/10.52324/001c.8060

National Center for Education Statistics. (n.d.). *Enrollment and percentage distribution of enrollment in public elementary and secondary schools, by race/ethnicity and level of education: Fall 1999 through fall 2031*. https://nces.ed.gov/programs/digest/d22/tables/dt22_203.60.asp

Peng, Y., Tian, C., & Wen, H. (2021). How does school district adjustment affect housing prices: An empirical investigation from Hangzhou, China. *China Economic Review*, *69*, 101683. https://doi.org/10.1016/j.chieco.2021.101683

Owusu-Edusei, K., Espey, M., & Lin, H. (2007). Does close count? school proximity, school quality, and residential property values. *Journal of Agricultural and Applied Economics*, *39*(1), 211–221. https://doi.org/10.1017/s1074070800022859

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34 - 55. https://doi.org/10.1086/260169

Rosiers, F. D., Laganà, A., & Thériault, M. (2001). Size and proximity effects of primary schools on surrounding house values. *Journal of Property Research*, *18*(2), 149–168. https://doi.org/10.1080/09599910110039905

Sah, V., Conroy, S. J., & Narwold, A. (2015). Estimating School Proximity Effects on House prices: the Importance of Robust Spatial Controls in Hedonic Estimations. *Journal of Real Estate Finance and Economics/the Journal of Real Estate Finance and Economics*, *53*(1), 50–76. https://doi.org/10.1007/s11146-015-9520-5

WPRDC. (2022). WPRDC. http://www.wprdc.org/about/

APPENDIX

APPENDIX A. Full Categorization of Allegheny County School Districts by Area Type

ALLEGHENY COUNTY SCHOOL DISTRICTS CATEGORIZED BY AREA TYPE				
#	School District	Urban	Suburban	Semi-rural
1	Allegheny Valley		✓	
2	Avonworth		✓	
3	Baldwin Whitehall		✓	
4	Bethel Park		✓	
5	Brentwood Borough		✓	
6	Carlynton		✓	
7	Chartiers Valley		✓	
8	Clairton City		✓	
9	Cornell		✓	
10	Deer Lakes			✓
11	Duquesne City		✓	
12	East Allegheny		✓	
13	Elizabeth Forward			✓
14	Fort Cherry			✓
15	Fox Chapel Area		✓	
16	Gateway		✓	
17	Hampton Township		✓	
18	Highlands			✓
19	Keystone Oaks		✓	
20	McKeesport Area		✓	
21	Montour		✓	
22	Moon Area		✓	
23	Mt Lebanon		✓	
24	North Allegheny		✓	
25	North Hills		✓	
26	Northgate		✓	
27	Penn Hills Twp		✓	
28	Penn-Trafford		✓	
29	Pine-Richland		✓	
30	Pittsburgh	✓		
31	Plum Borough		✓	
32	Quaker Valley			✓
33	Riverview		✓	
34	Shaler Area		✓	
35	South Allegheny		✓	
36	South Fayette Twp			✓
37	South Park		✓	
38	Steel Valley		✓	
39	Sto-Rox		✓	

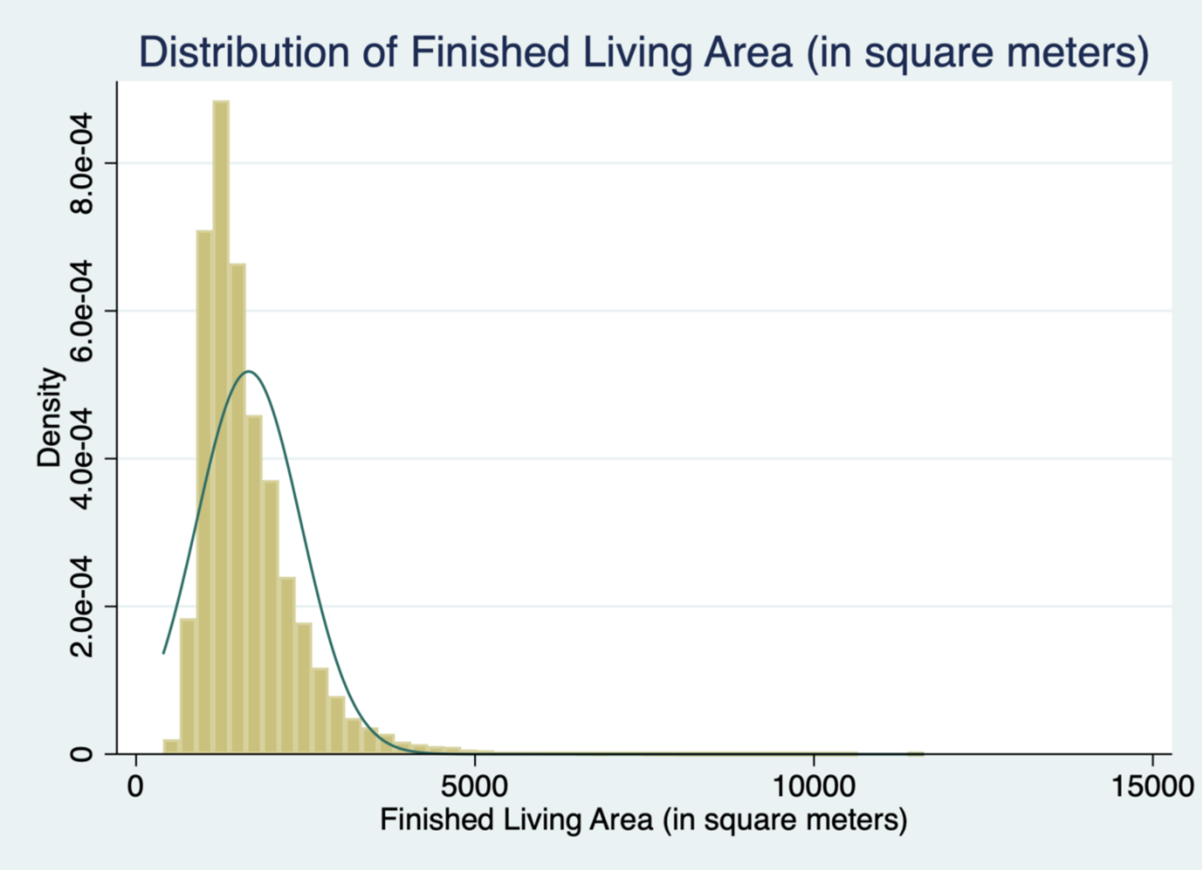
40	Upper St Clair		✓	
41	West Allegheny			✓
42	West Jefferson Hills		✓	
43	West Mifflin		✓	
44	Wilkinsburg Borough		✓	
45	Woodland Hills		✓	

APPENDIX B: Correlation matrix

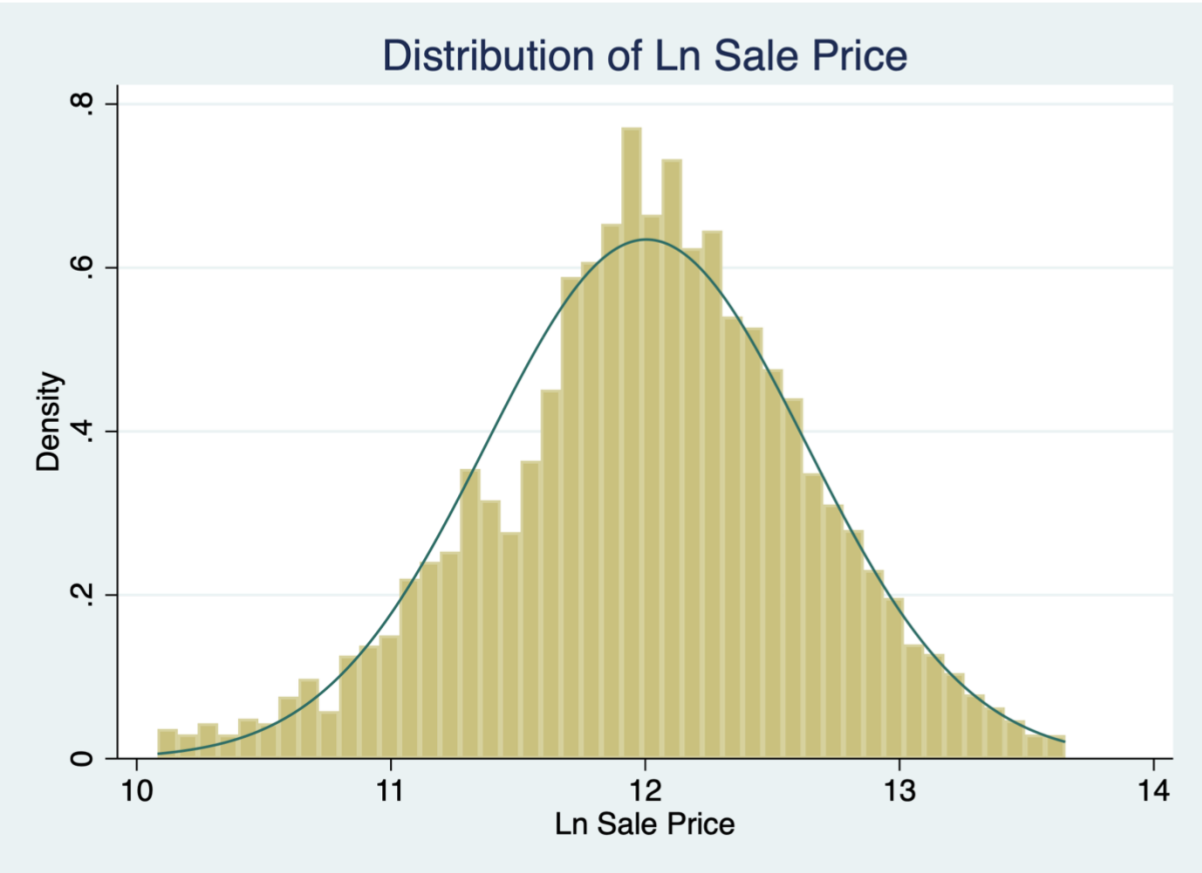
	FL Area	Property Age	# Bedrooms	Dummy AC	Dummy Basement	Dummy Apartment
FL Area	1.0000					
Property Age	- 0.1136	1.0000				
# Bedrooms	0.6253	- 0.0609	1.0000			
Dummy AC	0.0724	- 0.4829	0.0433	1.0000		
Dummy Basement	0.1207	0.1719	0.2692	- 0.0878	1.0000	
Dummy Apartment	- 0.1349	- 0.2250	- 0.2923	0.1393	- 0.6708	1.0000

APPENDIX C: Histrogram showing the distribution of the variable ‘Sale Price’

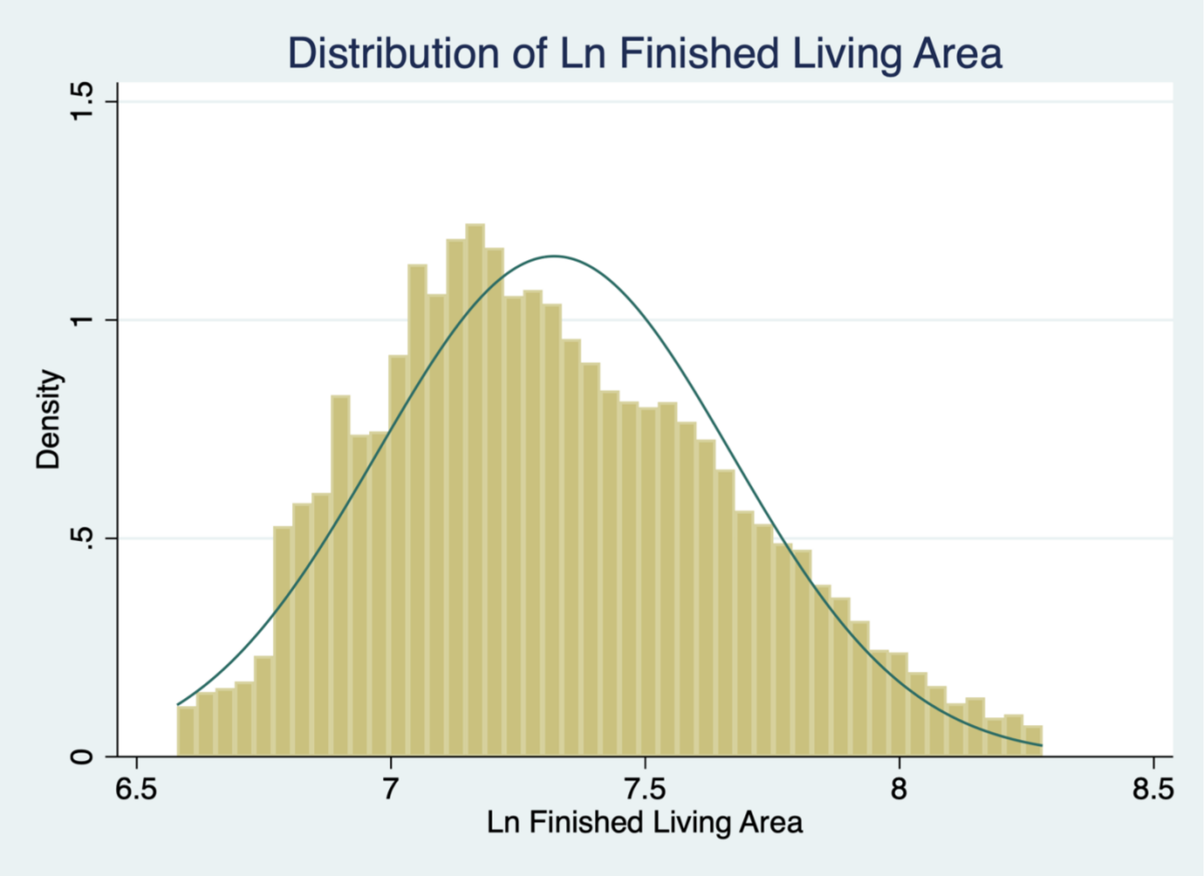
APPENDIX D: Histrogram showing the distribution of the variable ‘Finished Living Area’



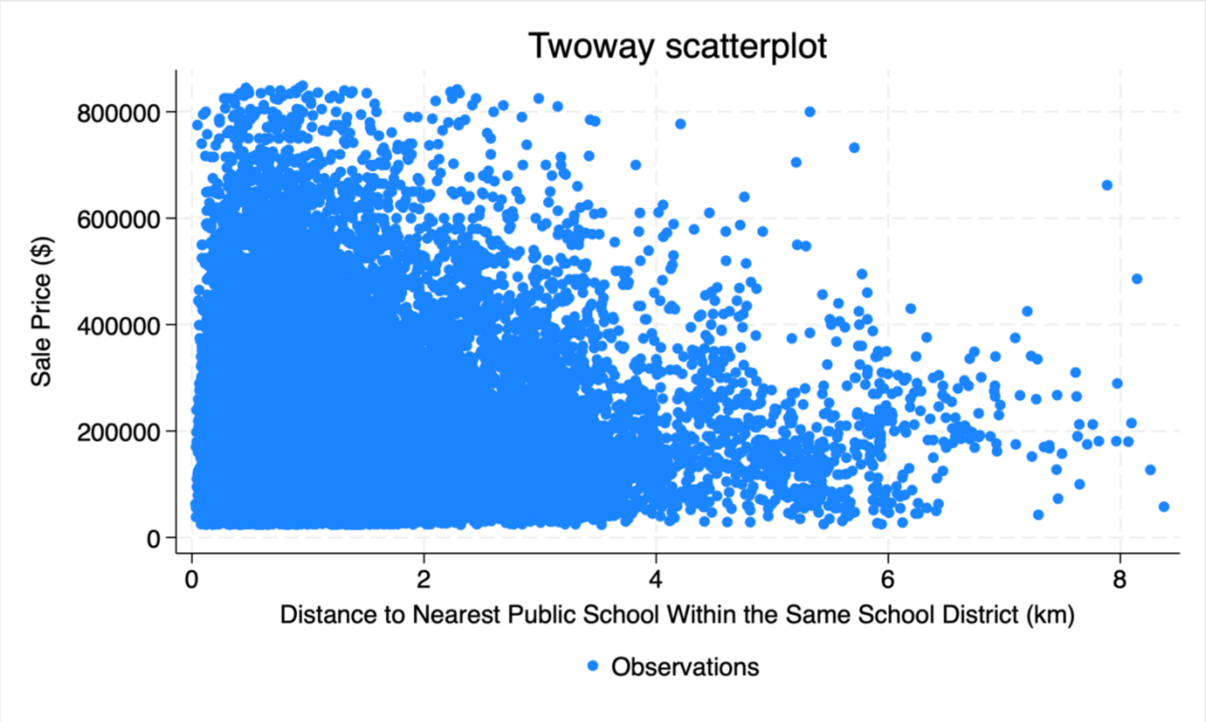
APPENDIX E: Histrogram showing the distribution of the variable ‘LN Sale Price’



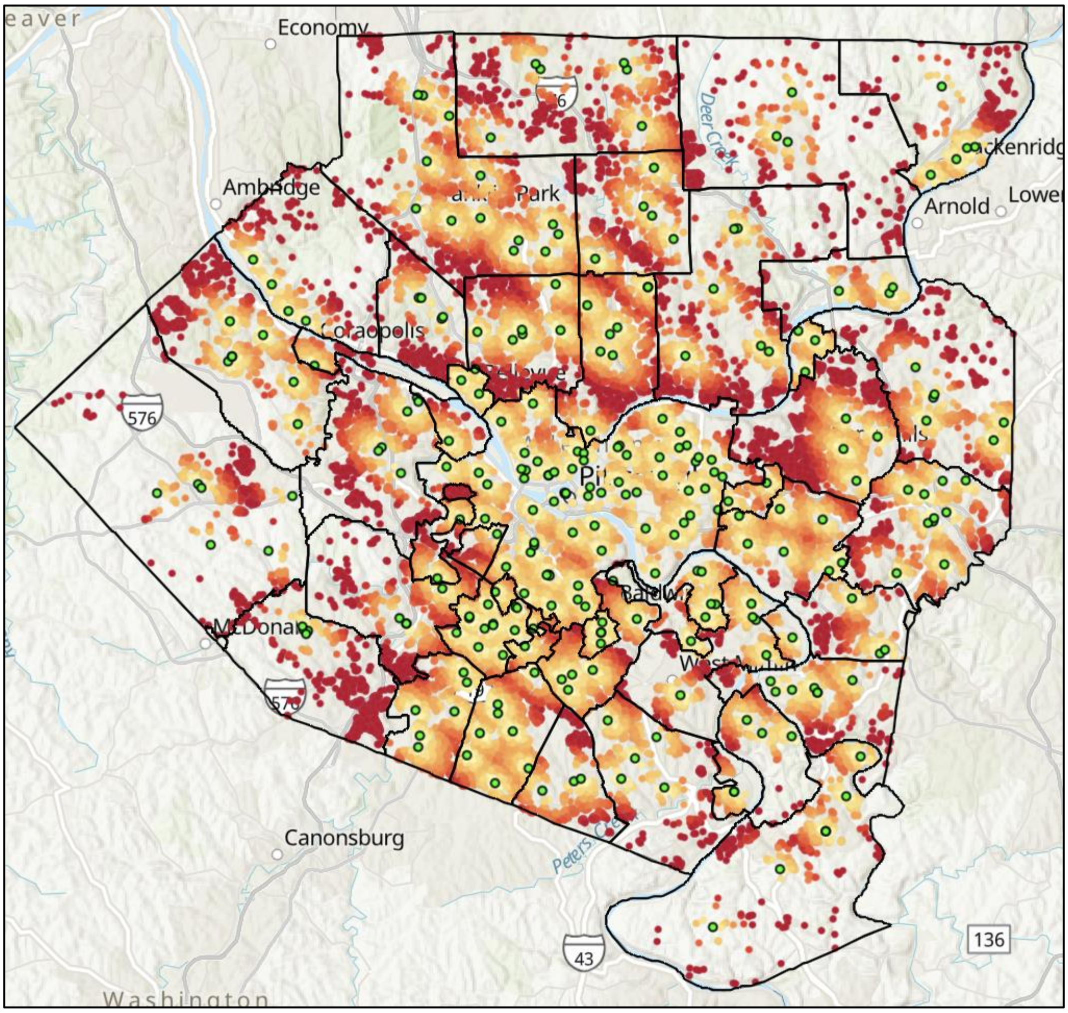
APPENDIX F: Histrogram showing the distribution of the variable 'LN Finished Living Area'



APPENDIX G: Twoway scatterplot



APPENDIX H: Zoomed-in image of map



APPENDIX I: Research data management plan

1. General	
1.1 Name & title of thesis	Proximity to Public Schools and House prices: Insights from Allegheny County, Pennsylvania
1.2 *(if applicable)* Organisation. Provide details on the organisation where the research takes place if this applies (in case of an internship).	

2 Data collection – the creation of data	
2.1. Which data formats or which sources are used in the project? For example: - theoretical research, using literature and publicly available resources	The datasets used in this study are obtained from the Western Pennsylvania Regional Data Center (WPRDC), an open-access database overseen by the University of Pittsburgh Center for Urban and Social

- Survey Data - Field Data - Interviews	Research, and funded jointly by Allegheny County and the City of Pittsburgh.
2.2 Methods of data collection What method(s) do you use for the collection of data. (Tick all boxes that apply)	☐ Structured individual interviews ☐ Semi-structured individual interviews ☐ Structured group interviews ☐ Semi-structured group interviews ☐ Observations ☐ Survey(s) ☐ Experiment(s) in real life (interventions) ☒ Secondary analyses on existing data sets (if so: please also fill in 2.3) ☐ Public sources (e.g. University Library) ☐ Other (explain):
2.3. (If applicable): if you have selected ‘Secondary analyses on existing datasets’: who provides the data set?	☐ Data is supplied by the University of Groningen. ☒ Data have been supplied by an external party. (Western Pennsylvania Regional Data Center)

3 Storage, Sharing and Archiving	
3.1 Where will the (raw) data be stored *during* research? If you want to store research data, it is good practice to ask yourself some questions: • How big is my dataset at the end of my research? • Do I want to collaborate on the data? • How confidential is my data? • How do I make sure I do not lose my data? Need more information? Take a look at the site of the Digital Competence Centre (DCC)) Feel free to contact the DCC for questions: dcc@rug.nl	☐ X-drive of UG network ☐ Y-drive of UG network ☐ (Shared) UG Google Drive ☐ Unishare ☒ Personal laptop or computer ☐ External devices (USB, harddisk, NAS) ☐ Other (explain):
3.2 Where are you planning to store / archive the data after you have finished your research? Please explain where and for how long. Also explain who has access to these data NB do not use a personal UG network or google drive for archiving data!	☐ X-drive of UG network ☐ Y-drive of UG network ☐ (Shared) UG Google Drive ☐ Unishare ☐ In a repository (i.e. DataverseNL) ☒ Other (personal laptop): The retention period will be [1] year.

3.3 Sharing of data With whom will you be sharing data during your research?	☒ University of Groningen ☐ Universities or other parties in Europe ☐ Universities or other parties outside Europe ☐ I will not be sharing data

4. Personal data	
4.1 Collecting personal data Will you be collecting personal data? If you are conducting research with personal data you have to comply to the General Data Privacy Regulation (GDPR). Please fill in the questions found in the appendix 3 on personal data.	No
If the answer to 4.1 is ‘no’, please skip the section below and proceed to section 5	
4.2 What kinds of categories of people are involved? Have you determined whether these people are vulnerable in any way (see FAQ)? If so, your supervisor will need to agree.	My research project involves: ☐ Adults (not vulnerable) ≥ 18 years ☐ Minors < 16 years ☐ Minors < 18 years ☐ Patients ☐ (other) vulnerable persons, namely (please provide an explanation what makes these persons vulnerable) (Please give a short description of the categories of research participants that you are going to involve in your research.)
4.3 Will participants be enlisted in the project without their knowledge and/or consent? (E.g., via covert observation of people in public places, or by using social media data.)	Yes/no If yes, please explain if, when and how you will inform the participants about the study.
4.4 Categories of personal data that are processed. Mention all types of data that you systematically collect and store. If you use particular kinds of software, then check what the software is doing as well. Of course, always ask yourself if you need all categories of data for your project.	☐ Name and address details ☐ Telephone number ☐ Email address ☐ Nationality ☐ IP-addresses and/or device type ☐ Job information ☐ Location data ☐ Race or ethnicity ☐ Political opinions

	☐ Physical or mental health ☐ Information about a person's sex life or sexual orientation ☐ Religious or philosophical beliefs ☐ Membership of a trade union ☐ Biometric information ☐ Genetic information ☐ Other (please explain below):
4.5 Technical/organisational measures Select which of the following security measures are used to protect personal data.	☐ Pseudonymisation ☐ Anonymisation ☐ File encryption ☐ Encryption of storage ☐ Encryption of transport device ☐ Restricted access rights ☐ VPN ☐ Regularly scheduled backups ☐ Physical locks (rooms, drawers/file cabinets) ☐ None of the above ☐ Other (describe below):
4.6 Will any personal data be transferred to organisations within countries outside the European Economic Area (EU, Norway, Iceland and Liechtenstein)? If the research takes places in a country outside the EU/EEA, then please also indicate this.	Yes/no If yes, please fill in the country.

5 – Final comments	
Do you have any other information about the research data that was not addressed in this template that you think is useful to mention?	No