Industrial Impact on Housing Prices in Western Kentucky

Brad Brauser

# Industrial Impact on Housing Prices in Western Kentucky

#### Brad Brauser

#### 6/25/2023

## Summary

## Introduction

In the western Kentucky area, the impact of proximity to industrial areas on property pricing is a significant concern. This project aims to examine how residential properties’ value is affected by their proximity to factories, packaging facilities, and other industries. By analyzing factors such as distance, pollution levels, and market dynamics, we seek to provide valuable insights for homeowners, real estate agents, and policymakers. Understanding the correlation between industrial proximity and property prices in this region is crucial for informed decision-making regarding real estate investments and urban planning.

## Literature Review

Although there is limited specific literature on the impact of industrial areas on property pricing in western Kentucky, studies conducted in other regions shed light on similar issues. One such study examined the property prices in Bordeaux, Dunkirk, and Rouen, which have hazardous chemical and petrochemical industries in or near the cities (Grislain-Letremy & Katossky, 2014). Another study focused on regions in the Netherlands, namely Rancid and Noord-Brabant (F. Vor & H. De Groot, 2011). Both papers found that the influence of industrial areas on property prices is predominantly localized, with the presence and size of industrial sites affecting the perceived spatial quality of the neighborhood. Furthermore, the willingness to pay for prevention measures varied significantly among different industrial areas (Grislain-Letremy & Katossky, 2014). While these studies offer valuable insights, there is a need for further research specifically addressing the western Kentucky area to understand the localized impact of industrial proximity on property pricing and the unique dynamics of the region.

Existing studies have examined the impact of industrial areas on property pricing in various regions, yet there remains a significant gap in research specifically addressing the western Kentucky area. This region possesses unique characteristics, including a distinct industrial landscape and residential patterns, which necessitate a dedicated study. Conducting research in this problem area will address the lack of localized understanding and insights into how proximity to industrial areas affects property values in western Kentucky. The study aims to investigate the direct correlation between distance from industrial sites and property pricing, the influence of pollution levels on housing market dynamics, and the varying perceptions of spatial quality in neighborhoods adjacent to different industrial sectors. By filling this research gap, the study will provide homeowners, real estate agents, and policymakers with essential information to make informed decisions regarding property investments, urban planning, and environmental management in the western Kentucky area. Ultimately, it will contribute to a comprehensive understanding of the impact of industrial proximity on property pricing and help shape strategies for sustainable development and equitable growth in the region.

Understanding the correlation between these factors is vital for homeowners, real estate agents, and policymakers who make decisions about property investments and urban planning. By investigating the proximity of residential areas to industrial sites, the study aims to reveal whether there is a direct correlation between distance and housing prices. This information will provide valuable insights into the dynamics of the housing market in Western KY, allowing stakeholders to make informed decisions based on the spatial relationship between industrial areas and property values. Additionally, the study’s findings will shed light on the potential impacts of industrial activities, such as pollution levels, on housing prices. This knowledge is essential for policymakers in developing strategies to ensure the sustainable development of the region while safeguarding the welfare of residents. By exploring these research questions, the study aims to contribute to evidence-based decision-making and support the development of a thriving real estate market in Western KY.

## Theory

This paper is exploring the impact between an industrial area’s location and the pricing surrounding residential properties. This exploration includes:

H1. The closer that a residential property is to an industrial area, the lower in price the property will be.

## Data

Housing data was downloaded in a CSV from <https://westky.paragonrels.com/ParagonLS/Default.mvc#1,1,2> and uploaded to <https://github.com/bbrauser/WKYIndustrialImpact/tree/main/Data>. The data includes:

# Specify the URL of the CSV file  
housing\_url <- "https://raw.githubusercontent.com/bbrauser/WKYIndustrialImpact/main/HousingResearch.csv"  
  
# Read the CSV file from the URL  
housing\_df <- read.csv(url(housing\_url))  
  
# Display the first 6 rows of the df  
head(housing\_df)

## Picture.Count MLS.. Price Class Type Status County  
## 1 4 118678 0 Residential Single Family Withdrawn Calloway  
## 2 1 99539 0 Residential Single Family Expired Ballard  
## 3 1 98723 550 Residential Single Family Sold Christian  
## 4 1 98726 550 Residential Single Family Sold Christian  
## 5 1 98725 1650 Residential Single Family Sold Christian  
## 6 34 118739 2000 Residential Single Family Rented Calloway  
## AddressNew Address  
## 1 2146 State Route 121 North, Murray, KY 2146 State Route 121 North  
## 2 229 Green Street, Wickliffe, KY 229 Green Street  
## 3 1892 Woodmill Road, Hopkinsville, KY 1892 Woodmill Road  
## 4 1910 Beach Street, Hopkinsville, KY 1910 Beach Street  
## 5 911 E 2nd Street, Hopkinsville, KY 911 E 2nd Street  
## 6 77 Wells Purdom Dr, Almo, KY 77 Wells Purdom Dr  
## City State Agent...Agent.Name Listing.Office.1...Office.Name  
## 1 Murray KY Shea Sykes SBG Real Property Professionals  
## 2 Wickliffe KY Mark Fredrick 1st Ky Realty and Auction  
## 3 Hopkinsville KY Michael Harris Harris Real Estate & Auction  
## 4 Hopkinsville KY Michael Harris Harris Real Estate & Auction  
## 5 Hopkinsville KY Michael Harris Harris Real Estate & Auction  
## 6 Almo KY Wilma Woods Loretta Jobs Realty  
## Latitude Longitude  
## 1 33.15950 -82.038204  
## 2 NA NA  
## 3 56.07060 -3.432663  
## 4 NA NA  
## 5 34.04503 -118.216583  
## 6 NA NA

Because some of these properties were either not sold or are rental properties, those rows can be removed. The following is a list of the values removed and the reason why they were removed:

Withdrawn: The property did not sell ContractReceived: The property is under contract, but the price it was closed for is unknown Expired: The property did not sell Rented: The property was a rental Auction: The property was auctioned off and not sold

# Define the values to be removed  
values\_to\_remove <- c("Withdrawn", "Contract Received", "Expired", "Rented", "Auction")  
  
# Remove the values from the Status column  
housing\_df <- subset(housing\_df, !(Status %in% values\_to\_remove))  
  
# Reset the index of the dataframe  
housing\_df <- data.frame(housing\_df, row.names = NULL)  
  
# Display the filtered dataframe  
head(housing\_df)

## Picture.Count MLS.. Price Class Type Status County  
## 1 1 98723 550 Residential Single Family Sold Christian  
## 2 1 98726 550 Residential Single Family Sold Christian  
## 3 1 98725 1650 Residential Single Family Sold Christian  
## 4 0 98733 2750 Residential Single Family Sold Christian  
## 5 1 98731 3190 Residential Single Family Sold Christian  
## 6 2 116526 4000 Residential Single Family Sold Caldwell  
## AddressNew Address City State  
## 1 1892 Woodmill Road, Hopkinsville, KY 1892 Woodmill Road Hopkinsville KY  
## 2 1910 Beach Street, Hopkinsville, KY 1910 Beach Street Hopkinsville KY  
## 3 911 E 2nd Street, Hopkinsville, KY 911 E 2nd Street Hopkinsville KY  
## 4 1112 E 18th Street, Hopkinsville, KY 1112 E 18th Street Hopkinsville KY  
## 5 317 E 17th Street, Hopkinsville, KY 317 E 17th Street Hopkinsville KY  
## 6 118 Baldwin, Princeton, KY 118 Baldwin Princeton KY  
## Agent...Agent.Name Listing.Office.1...Office.Name  
## 1 Michael Harris Harris Real Estate & Auction  
## 2 Michael Harris Harris Real Estate & Auction  
## 3 Michael Harris Harris Real Estate & Auction  
## 4 Michael Harris Harris Real Estate & Auction  
## 5 Michael Harris Harris Real Estate & Auction  
## 6 Brandon K Warfield Keller Williams Experience Realty Eddyville Branch  
## Latitude Longitude  
## 1 56.07060 -3.432663  
## 2 NA NA  
## 3 34.04503 -118.216583  
## 4 32.67014 -117.092007  
## 5 40.73418 -73.982687  
## 6 NA NA

Some rows did not have values in Latitude and Longitude, so those were removed:

# Remove rows with blank cells in Latitude and Longitude columns  
housing\_df <- housing\_df[complete.cases(housing\_df$Latitude, housing\_df$Longitude), ]  
  
# Display the filtered dataframe  
head(housing\_df)

## Picture.Count MLS.. Price Class Type Status County  
## 1 1 98723 550 Residential Single Family Sold Christian  
## 3 1 98725 1650 Residential Single Family Sold Christian  
## 4 0 98733 2750 Residential Single Family Sold Christian  
## 5 1 98731 3190 Residential Single Family Sold Christian  
## 8 1 99782 4400 Residential Single Family Sold Christian  
## 10 5 120796 5250 Residential Single Family Sold Christian  
## AddressNew Address City State  
## 1 1892 Woodmill Road, Hopkinsville, KY 1892 Woodmill Road Hopkinsville KY  
## 3 911 E 2nd Street, Hopkinsville, KY 911 E 2nd Street Hopkinsville KY  
## 4 1112 E 18th Street, Hopkinsville, KY 1112 E 18th Street Hopkinsville KY  
## 5 317 E 17th Street, Hopkinsville, KY 317 E 17th Street Hopkinsville KY  
## 8 1122 YOUNGLOVE ST, Hopkinsville, KY 1122 YOUNGLOVE ST Hopkinsville KY  
## 10 2007 Carl Smith Rd, Hopkinsville, KY 2007 Carl Smith Rd Hopkinsville KY  
## Agent...Agent.Name Listing.Office.1...Office.Name Latitude  
## 1 Michael Harris Harris Real Estate & Auction 56.07060  
## 3 Michael Harris Harris Real Estate & Auction 34.04503  
## 4 Michael Harris Harris Real Estate & Auction 32.67014  
## 5 Michael Harris Harris Real Estate & Auction 40.73418  
## 8 Joshua D Abner Atlas Real Estate & Auction Services 36.86631  
## 10 Shulorn Hollowell Jeter eXp Realty LLC 36.87231  
## Longitude  
## 1 -3.432663  
## 3 -118.216583  
## 4 -117.092007  
## 5 -73.982687  
## 8 -87.473495  
## 10 -87.465573

To give more of a clear minimum price, the Price column was filtered to show only prices of $20,000 and above:

# Filter the price column to show values 20,000 and above  
housing\_df <- filter(housing\_df, Price >= 20000)  
  
# Display the first 6 rows of the df  
head(housing\_df)

## Picture.Count MLS.. Price Class Type Status County  
## 1 1 84937 20000 Residential Single Family Sold Ballard  
## 2 1 81066 20000 Residential Single Family Sold Caldwell  
## 3 1 99770 20900 Residential Single Family Sold Christian  
## 4 1 106720 21000 Residential Single Family Sold Carlisle  
## 5 1 85498 21200 Residential Single Family Sold Caldwell  
## 6 1 88661 22000 Residential Single Family Sold Calloway  
## AddressNew Address  
## 1 5192 Holloway Landing Road, Barlow, KY 5192 Holloway Landing Road  
## 2 620 N Jeffereson, Princeton, KY 620 N Jeffereson  
## 3 420 East 23rd Street, Hopkinsville, KY 420 East 23rd Street  
## 4 204 Hobbs Street, Arlington, KY 204 Hobbs Street  
## 5 3008 Grooms Lane, Princeton, KY 3008 Grooms Lane  
## 6 400 N 7th Street, Murray, KY 400 N 7th Street  
## City State Agent...Agent.Name Listing.Office.1...Office.Name  
## 1 Barlow KY Kimberly M Spell Purchase Realty Group  
## 2 Princeton KY Billie K Farless Farless Realty  
## 3 Hopkinsville KY Joshua D Abner Atlas Real Estate & Auction Services  
## 4 Arlington KY Deana C O'Neal Bone & Co. Realtors  
## 5 Princeton KY Jesse Case RE/MAX Real Estate Services  
## 6 Murray KY Gale Broach Sharp Kopperud Realty  
## Latitude Longitude  
## 1 37.05255 -89.06591  
## 2 41.89015 -87.64302  
## 3 40.73601 -73.97764  
## 4 -37.80792 144.89303  
## 5 39.32841 -76.87252  
## 6 39.95827 -75.15042

Some of the data was not needed, so the housing data was limited to Address, County, Price, Latitude, and Longitude:

# Removes unneeded columns from df  
housing\_df <- select(housing\_df, Address, County, Price, Latitude, Longitude)  
  
# Display the first 6 rows of the df  
head(housing\_df)

## Address County Price Latitude Longitude  
## 1 5192 Holloway Landing Road Ballard 20000 37.05255 -89.06591  
## 2 620 N Jeffereson Caldwell 20000 41.89015 -87.64302  
## 3 420 East 23rd Street Christian 20900 40.73601 -73.97764  
## 4 204 Hobbs Street Carlisle 21000 -37.80792 144.89303  
## 5 3008 Grooms Lane Caldwell 21200 39.32841 -76.87252  
## 6 400 N 7th Street Calloway 22000 39.95827 -75.15042

Location information of the industrial areas in western Kentucky was downloaded in a CSV from <https://www.irs.gov/pub/irs-soi/17incyky.xlsx> and uplaoded to GitHub. The data includes:

# Specify the URL of the CSV file  
industry\_url <- "https://raw.githubusercontent.com/bbrauser/WKYIndustrialImpact/main/IndustrialAreas.csv"  
  
# Read the CSV file from the URL  
industry\_df <- read.csv(url(industry\_url))  
  
# Display the first 6 rows of the df  
head(industry\_df)

## X Y OBJECTID\_1 OBJECTID CompanyID County  
## 1 -9525225 4609369 1 1 001t0000003slbi Jefferson  
## 2 -9835992 4420672 2 2 001t0000003sjBS Marshall  
## 3 -9567116 4439758 3 3 001t0000003skTh Barren  
## 4 -9538261 4608138 4 4 001t0000003sn4S Jefferson  
## 5 -9383639 4632518 5 5 001t0000003sjEV Harrison  
## 6 -9425807 4733197 6 6 001t0000003sk1X Boone  
## FacilityNa  
## 1 310 Tempering  
## 2 3A Composites USA Inc  
## 3 3A Composites USA Inc  
## 4 3DR Laboratories, LLC  
## 5 3M  
## 6 3M Hebron  
## Address  
## 1 2409 Plantside Drive, Louisville, KY 40299 United States  
## 2 208 W 5th St, Benton, KY 42025-1100 United States  
## 3 205 American Ave, Glasgow, KY 42141-1136 United States  
## 4 1941 Bishop Lane, Suite 807, Louisville, KY 40218 United States  
## 5 1308 New Lair Rd, Cynthiana, KY 41031-8823 United States  
## 6 1151 Aviation Blvd, Hebron, KY 41048-9333 United States  
## ProductsSe  
## 1 310 Tempering specializes in precise fabrication and tempering for applications such as heavy glass shower enclosures, railings, interior partitions, all glass entrances, table tops and much more.  
## 2 Composite aluminum & plastic foam panels & sheets  
## 3 Paper faced foam panels, thick foam panels and graphic arts boards\nPET Foam for the core material market  
## 4 Radiology information technology service company that allows diagnostic imaging facilities to outsource their 3D image processing. By outsourcing all or part of this function, clients reduce cost, improve quality and can operate 24/7.  
## 5 Office supplies and stationery products: repositionable note pads, easel pads, and packaging tapes. Post-it (R) Notes.  
## 6 Compound custom resins & specialty thermoplastic resin pellets  
## Employees YrEst Latitude Longitude ImportKBIF CompanyNam  
## 1 78 2015 38.21165 -85.56654 569113104 310 Tempering  
## 2 130 1978 36.86762 -88.35822 1091 3A Composites USA Inc  
## 3 142 1993 37.00467 -85.94286 4678 3A Composites USA Inc  
## 4 130 2006 38.20296 -85.68365 297407900 3DR Labs II, LLC  
## 5 550 1969 38.37486 -84.29466 1304 3M Company  
## 6 43 1995 39.08041 -84.67346 3312 3M  
## Phone Website ByProducts  
## 1 (502) 749-4141 www.310tempering.com   
## 2 (270) 527-4200 www.3acompositesusa.com   
## 3 270-651-3822 www.graphicdisplayusa.com   
## 4 (502) 814-7018 www.3drinc.com   
## 5 859-234-5671 www.3M.com   
## 6 859-334-4500 www.3m.com 001t0000003sk1XAAQ  
## ModDate Mfg NAICSCode1 NAICSCode2 NAICSCode3 NAICSCode4  
## 1 2021/11/11 00:00:00+00 1 327215 NA NA NA  
## 2 2021/11/29 00:00:00+00 1 331318 326150 326113 NA  
## 3 2022/02/18 00:00:00+00 1 326140 NA NA NA  
## 4 2021/11/11 00:00:00+00 0 621512 NA NA NA  
## 5 2022/02/21 00:00:00+00 1 322220 322121 NA NA  
## 6 2021/10/13 00:00:00+00 1 325211 NA NA NA  
## NAICSCode5 NAICSCode6 Main\_NAICS NAICS6  
## 1 NA NA 327215 327215  
## 2 NA NA 331318 331318  
## 3 NA NA 326140 326140  
## 4 NA NA 621512 621512  
## 5 NA NA 322220 322220  
## 6 NA NA 325211 325211  
## NAICSNatlI  
## 1 Glass Product Manufacturing Made of Purchased Glass  
## 2 Other Aluminum Rolling, Drawing and Extruding  
## 3 Polystyrene Foam Product Manufacturing  
## 4 Diagnostic Imaging Centers  
## 5 Paper Bag and Coated and Treated Paper Manufacturing  
## 6 Plastics Material and Resin Manufacturing  
## NAICSNat\_1 fPotMfgAsp YrRmvd ESRI\_OID  
## 1 Glass Product Mfg. Made of Purchased Glass 0 0 296  
## 2 Other Aluminum Rolling, Drawing and Extruding 0 0 316  
## 3 Polystyrene Foam Product Manufacturing 0 0 281  
## 4 Diagnostic Imaging Centers 0 0 925  
## 5 Paper Bag and Coated and Treated Paper Mfg 0 0 233  
## 6 Plastics Material and Resin Manufacturing 0 0 253

This data only needs County, FacilityNa, Longitude, and Latitude:

# Compiles dataframe from industry\_df  
industry\_df <- select(industry\_df, County, FacilityNa, Longitude, Latitude)  
  
# Display the first 6 rows of the df  
head(industry\_df)

## County FacilityNa Longitude Latitude  
## 1 Jefferson 310 Tempering -85.56654 38.21165  
## 2 Marshall 3A Composites USA Inc -88.35822 36.86762  
## 3 Barren 3A Composites USA Inc -85.94286 37.00467  
## 4 Jefferson 3DR Laboratories, LLC -85.68365 38.20296  
## 5 Harrison 3M -84.29466 38.37486  
## 6 Boone 3M Hebron -84.67346 39.08041

Since this df includes industrial locations in the whole state of Kentucky, the data must be filtered to include only the counties that are present in the houisng data CSV:

industry\_df <- industry\_df[industry\_df$County %in% housing\_df$County, ]  
  
# Resets the index  
industry\_df <- data.frame(industry\_df, row.names = NULL)  
  
# Display the first 6 rows of the df  
head(industry\_df)

## County FacilityNa Longitude Latitude  
## 1 Marshall 3A Composites USA Inc -88.35822 36.86762  
## 2 Graves A.C.E. Compressor Parts & Services Inc -88.65970 36.76000  
## 3 McCracken ACBL -88.72713 37.05047  
## 4 McCracken Acoustic Design Inc -88.60853 37.08665  
## 5 Lyon Acuren Inspection Inc. -88.13272 37.06385  
## 6 Caldwell AEK Transport Services LLC -87.88230 37.11703

## Methodology

After cleaning and preparing the data for analysis, a new dataframe (‘price\_analysis\_df’) was created that concatenated

# Initialize an empty data frame to store the results  
price\_analysis\_df <- data.frame(Address = housing\_df$Address,   
 Price = housing\_df$Price,   
 stringsAsFactors = FALSE)   
  
# Add columns for closest industry and distance   
price\_analysis\_df$Industry <- ""   
price\_analysis\_df$Distance <- Inf   
  
# Calculate distance between each property and industry   
for (property\_row in 1:nrow(housing\_df)) {  
 property\_lat <- housing\_df$Latitude[property\_row]  
 property\_lon <- housing\_df$Longitude[property\_row]  
   
 for (industry\_row in 1:nrow(industry\_df)) {  
 industry\_lat <- industry\_df$Latitude[industry\_row]  
 industry\_lon <- industry\_df$Longitude[industry\_row]  
   
 distance <- round(sqrt((property\_lat - industry\_lat)^2 + (property\_lon - industry\_lon)^2), 2)  
   
 # Check if the current distance is smaller than the previously recorded minimum distance  
 if (distance < price\_analysis\_df$Distance[property\_row]) {   
 price\_analysis\_df$Industry[property\_row] <- industry\_df$FacilityNa[industry\_row]   
 price\_analysis\_df$Distance[property\_row] <- distance  
 }  
 }  
}   
  
# Print the resulting data frame   
head(price\_analysis\_df)

## Address Price Industry Distance  
## 1 5192 Holloway Landing Road 20000 Harris Engineering Inc 0.03  
## 2 620 N Jeffereson 20000 Little Kentucky Smokehouse 4.11  
## 3 420 East 23rd Street 20900 CWP Industries 13.75  
## 4 204 Hobbs Street 21000 Koppers Inc 243.70  
## 5 3008 Grooms Lane 21200 CWP Industries 10.59  
## 6 400 N 7th Street 22000 CWP Industries 12.42

Because some of these properties are not very close to an industry, we are going to filter out this dataframe to only include properties that are 20 miles away or less from an industry in industry\_df. This still gives us about 8000 properties to analyze:

# Filter the price column to show values 20,000 and above  
price\_analysis\_df <- filter(price\_analysis\_df, Distance <= 20)  
  
# Display the first 6 rows of the df  
head(price\_analysis\_df)

## Address Price Industry Distance  
## 1 5192 Holloway Landing Road 20000 Harris Engineering Inc 0.03  
## 2 620 N Jeffereson 20000 Little Kentucky Smokehouse 4.11  
## 3 420 East 23rd Street 20900 CWP Industries 13.75  
## 4 3008 Grooms Lane 21200 CWP Industries 10.59  
## 5 400 N 7th Street 22000 CWP Industries 12.42  
## 6 606 Westvaco Road 23000 Stella-Jones Corporation 1.42

Below is a simple scatterplot that shows the relationship between Price and Distance:

# Get the maximum values   
max\_price <- max(price\_analysis\_df$Price)  
max\_dist <- max(price\_analysis\_df$Distance)  
  
# Scatterplot of Price vs Distance  
ggplot(price\_analysis\_df, aes(x = Price, y = Distance)) +   
 geom\_point(color = "blue", size = 3) +   
 geom\_smooth(method = "lm", se = FALSE) +   
 labs(  
 title = "Price vs Distance",   
 x = "Price (USD)",   
 y = "Distance (miles)"   
 ) +  
 theme\_bw() +   
 theme(  
 plot.title = element\_text(size = 15, face = "bold"),   
 axis.title = element\_text(size = 12)   
 ) +  
 scale\_x\_continuous(breaks = seq(0, max\_price, 75000), limits = c(0, max\_price), labels = comma) +   
 scale\_y\_continuous(breaks = seq(0, max\_dist, 1), limits = c(0, max\_dist)) +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 28 rows containing missing values (geom\_smooth).

![](data:image/png;base64;base64,)

The residual plot below shows if the residuals are randomly distributed, which is an assumption of linear regression:

# Fit a linear model   
lm <- lm(Price ~ Distance, data = price\_analysis\_df)  
  
# Extract the residuals   
residuals <- residuals(lm)  
  
# Get the maximum values   
max\_price <- max(price\_analysis\_df$Price)  
max\_dist <- max(price\_analysis\_df$Distance)  
  
# Scatterplot of Price vs Distance, colored by residuals   
ggplot(price\_analysis\_df, aes(x = Price, y = Distance, color = residuals)) +   
 geom\_point(size = 3) +   
 geom\_smooth(method = "lm", se = FALSE) +   
 scale\_color\_gradient2(low = "blue", high = "red", mid = "white") +  
 labs(  
 title = "Price vs Distance",   
 subtitle = "Colored by Residuals",  
 x = "Price (USD)",   
 y = "Distance (miles)"   
 ) +  
 theme\_bw() +   
 theme(  
 plot.title = element\_text(size = 15, face = "bold"),  
 plot.subtitle = element\_text(size = 12, face = "italic"),  
 axis.title = element\_text(size = 12),  
 legend.title = element\_text(size = 12)  
 ) +  
 scale\_x\_continuous(breaks = seq(0, max\_price, 75000), limits = c(0, max\_price), labels = comma) +   
 scale\_y\_continuous(breaks = seq(0, max\_dist, 1), limits = c(0, max\_dist)) +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 28 rows containing missing values (geom\_smooth).

![](data:image/png;base64;base64,)

The quantile-quantile (or QQ) plot below shows the quantiles of the residuals to the quantiles of a normal distribution, allowing one to check if the residuals are normally distributed, which is another assumption of linear regression. Deviations from the diagonal line indicate non-normality.

# Create the QQ plot  
ggplot(price\_analysis\_df, aes(sample = Distance)) +  
 stat\_qq() +  
 geom\_abline(color = "red") +  
 xlab("Theoretical Quantiles") +  
 ylab("Sample Quantiles") +  
 ggtitle("QQ Plot of Distance")

![](data:image/png;base64;base64,)

ggplot(price\_analysis\_df, aes(sample = Price)) +  
 stat\_qq() +  
 geom\_abline(color = "red") +  
 xlab("Theoretical Quantiles") +  
 ylab("Sample Quantiles") +  
 ggtitle("QQ Plot of Price")

![](data:image/png;base64;base64,)

## Results

The plots shown above give a visual representation of the relationship between Price and Distance which seems to indicate that there is a negative relationship between Price and Distance. That is, when Distance increases, Price decreases. To give a more granualr look at the relationship, a regressions table was created:

# Fit linear regression model  
model <- lm(Price ~ Distance, data = price\_analysis\_df)  
  
# Create regression table  
regression\_table <- stargazer(model,   
 title = "Regression Results",  
 dep.var.labels = "Price",  
 covariate.labels = "Distance",  
 header = FALSE,  
 type = "text",  
 digits = 3)

##   
## Regression Results  
## ===============================================  
## Dependent variable:   
## ---------------------------  
## Price   
## -----------------------------------------------  
## Distance -1,753.965\*\*\*   
## (225.344)   
##   
## Constant 170,811.700\*\*\*   
## (1,663.992)   
##   
## -----------------------------------------------  
## Observations 8,734   
## R2 0.007   
## Adjusted R2 0.007   
## Residual Std. Error 113,127.900 (df = 8732)   
## F Statistic 60.583\*\*\* (df = 1; 8732)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# Print regression table  
cat(regression\_table, sep = "\n")

##   
## Regression Results  
## ===============================================  
## Dependent variable:   
## ---------------------------  
## Price   
## -----------------------------------------------  
## Distance -1,753.965\*\*\*   
## (225.344)   
##   
## Constant 170,811.700\*\*\*   
## (1,663.992)   
##   
## -----------------------------------------------  
## Observations 8,734   
## R2 0.007   
## Adjusted R2 0.007   
## Residual Std. Error 113,127.900 (df = 8732)   
## F Statistic 60.583\*\*\* (df = 1; 8732)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

The results indicated a significant negative effect of Distance on Price (-1,753.965, p < 0.01), suggesting that as the distance increased, the price decreased. However, there was some uncertainty in the estimate (standard error: 225.344).

The constant term in the regression equation represented 170,811.700 (p < 0.01), which was the expected price when the distance was zero. The R-squared value was 0.007, indicating that only 0.7% of the variation in Price was explained by Distance. The model’s predictive power was limited, as seen from the low R-squared value. The residual standard error was 113,127.900, indicating the average prediction error.

Overall, the findings suggested that Distance had a significant but weak influence on Price.

## Conclusion and Implications

The initial theory that the closer to an industrial area a property is the more negatively the price is affected. This was proven not to be the case, as the closer a property was to an industrial area, the higher the price of the property was. While this study was limited to the western KY area, it might be beneficial to see the implications of not only the entire state of Kentucky, but also how the type of industry affects the price.

# References

de Vor, F., & de Groot, H. L. (2011). The impact of industrial sites on residential property values: A hedonic pricing analysis from the Netherlands. Regional Studies, 45(5), 609–623. <https://doi.org/10.1080/00343401003601925>

Grislain-Letrrmy, Cc., & Katossky, A. (2014). The impact of hazardous industrial facilities on housing prices: A comparison of parametric and semiparametric hedonic price models. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.2457195>

Team kentucky. Team Kentucky | Cabinet for Economic Development. (n.d.). <https://ced.ky.gov/> existing\_industries/Major\_Industries