The Effects of City Noise on Real Estate

Walter Gray

5/12/2021

Abstract

This report addresses the question of whether city noise has an effect on housing prices. My goal was to understand the magnitude of the effect by applying a random forest model against the data. The analysis found a modest association between noise and home value. The relationship was expected to be linear but wound up being more complex: house prices increase as noise decreases up to a certain point, but then prices fall as noise drops further.

Introduction

Noise pollution represents a serious risk to public health and well being. The growing body of research about the negative impacts of noise pollution has filtered into local and national policy debates about sound ordinances, the costs of noise pollution, and mitigation techniques.^1 ^2

Realtors present prospective homebuyers with lists of amenities and pictures but understanding the ambient noise of a house is difficult unless visited in person. In addition, intermittent noises like trains, planes, or nearby fire stations may not be fully comprehended during a brief visit.

The role of traffic and other noise on real estate isn't yet thoroughly understood on a wide scale. Some real estate websites (like Realtor.com) are beginning to incorporate noise estimates but most of the larger industry leaders have yet to adopt the practice. This project looks to better understand how home values are affected by noise.

Method

Real estate data for Austin wasn't readily available in any open datasets, so I used Mechanical Turk to build a dataset from Redfin of all of the homes sold in the Cherrywood neighborhood of Austin over the last five years.

After building the dataset with some of the relevant features included, I enriched the addresses with latitude and longitude data obtained from latlong.net.

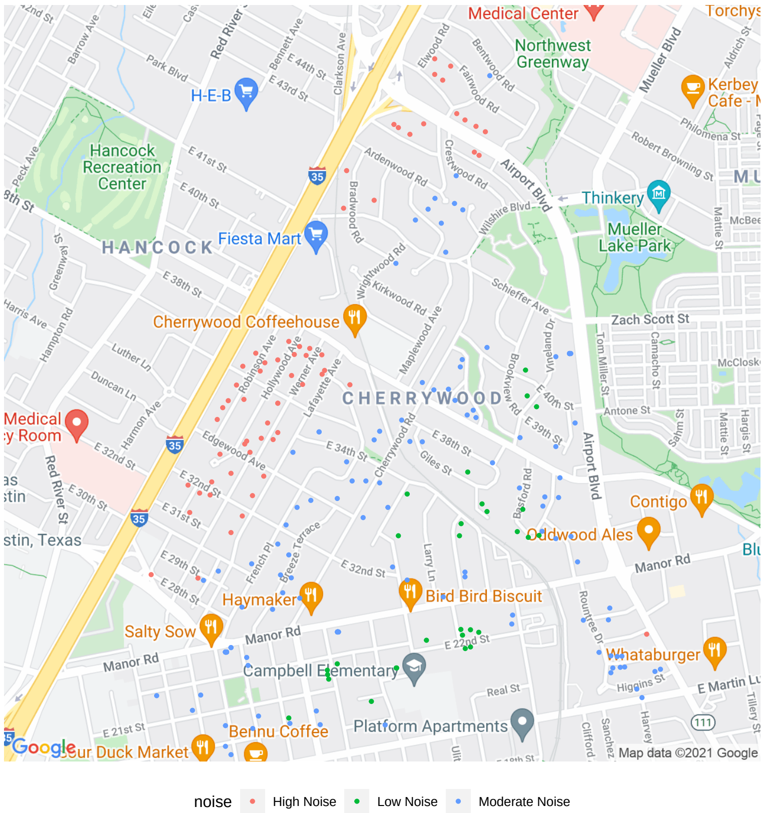
Sound data was brought in from howloud.com, a data vendor used by Realtor.com and several local MLS databases to estimate the relative noise likely to be experienced by a given property. Their "Soundscore" rating is a numeric score comprised from street traffic (gleaned from the Federal Highway Authority's Traffic Noise Model), overhead flights, and an estimate tied to local points of interest such as restaurants and schools. Lower scores represent louder environments than higher scores.

Several new features were created for the dataset: "month" and "year" were derived from "date_sold" and the categorical variables in "noise" were applied to the tertiles of the "soundscore" feature.

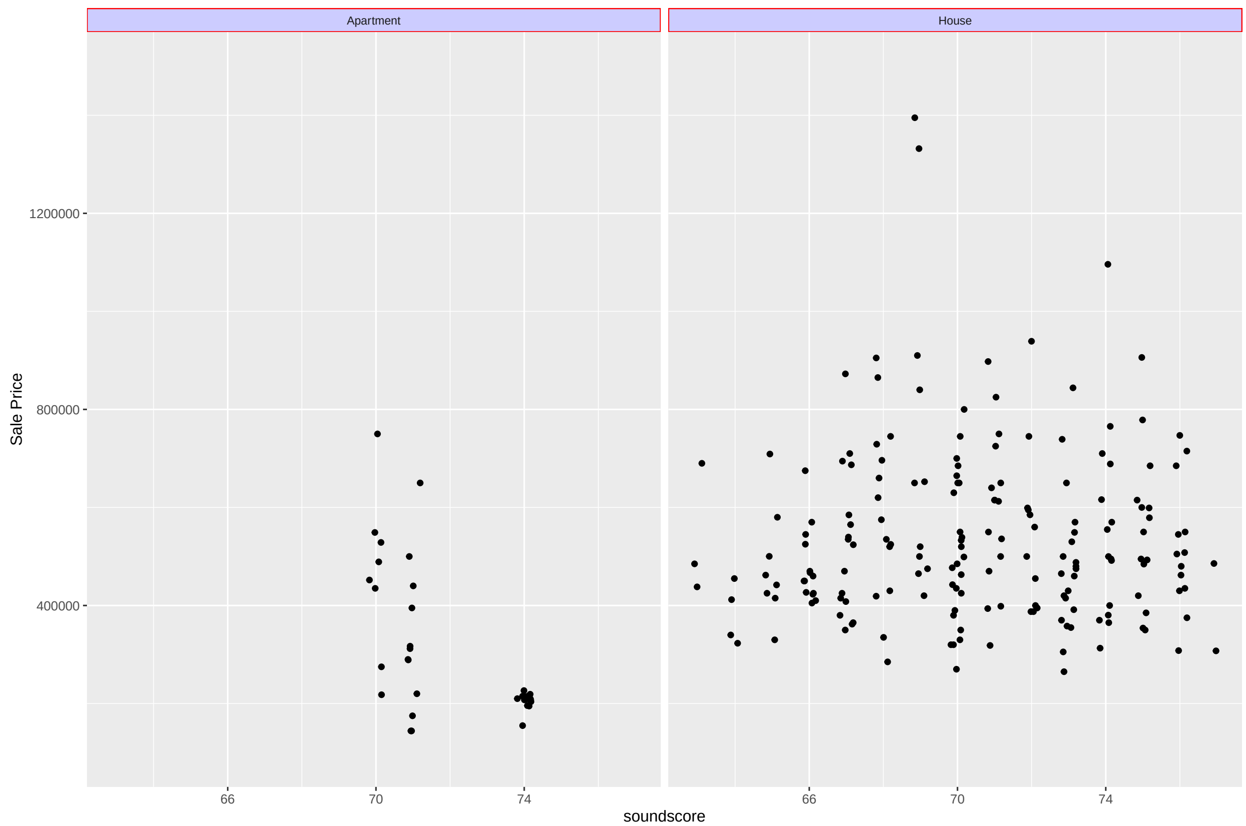
With the data appended with the noise scores and lat/long, I used ggmap to plot the houses and apartments on Google Maps. This exploratory data analysis was taken further by plotting the data with a jittered scatterplot and a boxplot.

After getting a sense for the problem at hand, I used a random forest algorithm as my primary analytical method combined with a partial dependence plot to understand the relationship between relative noise and sale price.

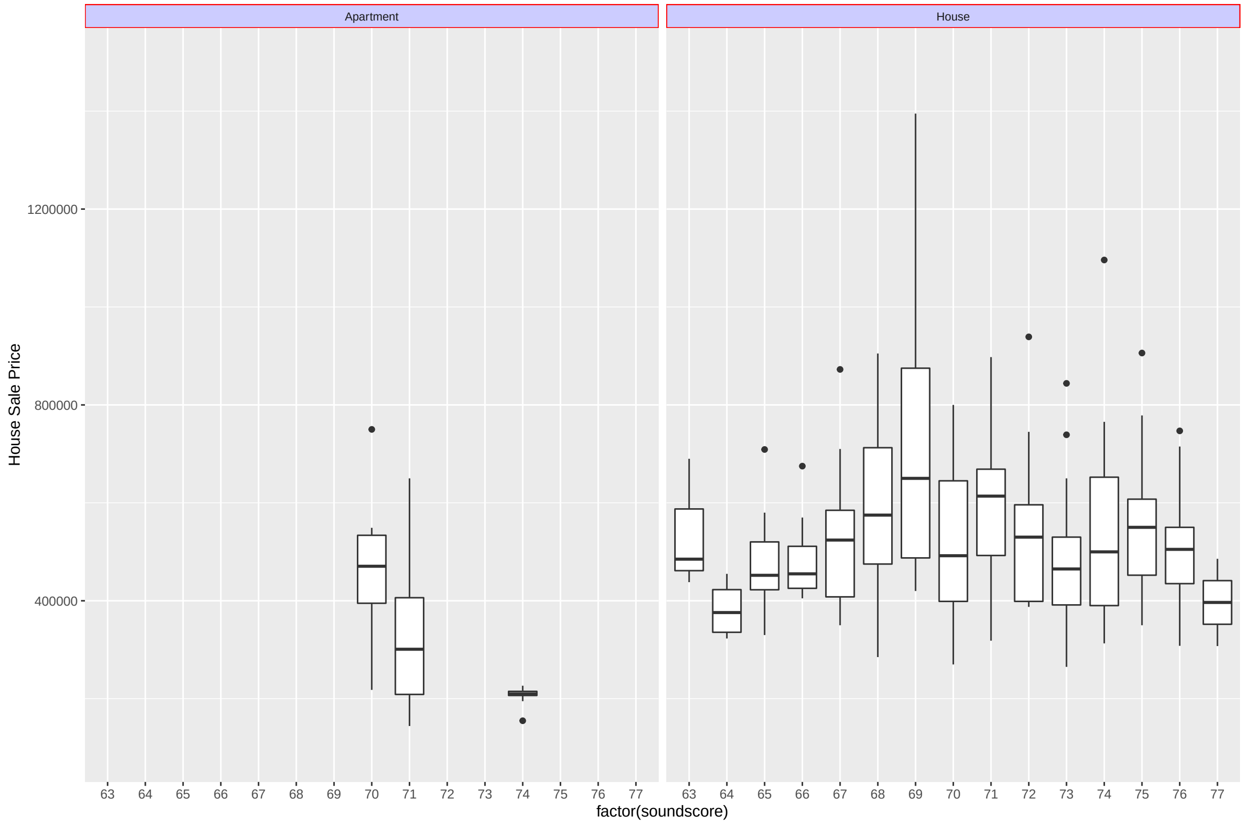
Results



This map of the Cherrywood neighborhood shows all of the homes sold in the area from 2016 to 2021 with color-coding for high, medium, and low noise levels. Many of the labels intuitively reflect proximity to the main highway that runs through the city on the western border of the neighborhood and the relatively high-thoroughfare streets of Airport and Manor that form the northern and southern borders of the neighborhood.



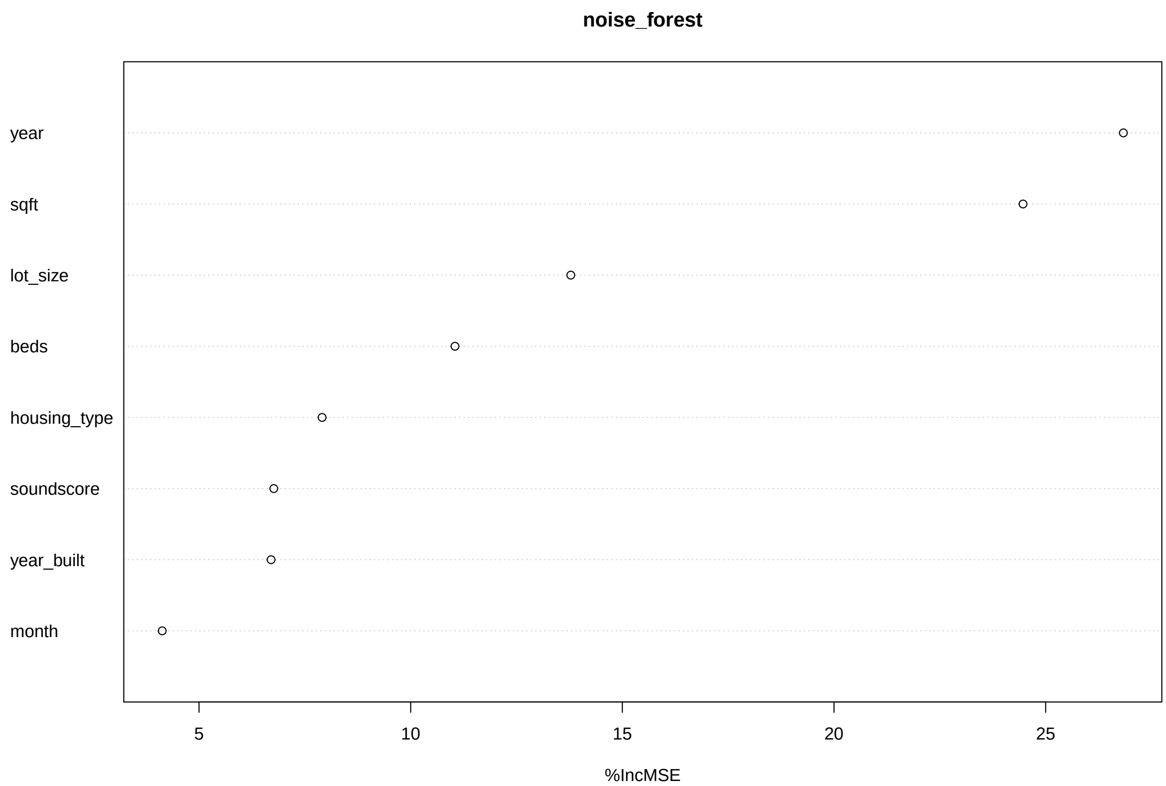
To get an introductory sense of the dataset I created a scatterplot faceted by housing type. Data is slightly jittered to communicate multiple housing units that score similarly on price and sound levels (to reiterate for clarity: a lower score reflects a louder environment, a higher score is quieter). House sales with noise levels at 66 or lower don't exceed $800,000. No visible clustering at other sound levels. Apartments show significant clustering at the sound rating of 74.



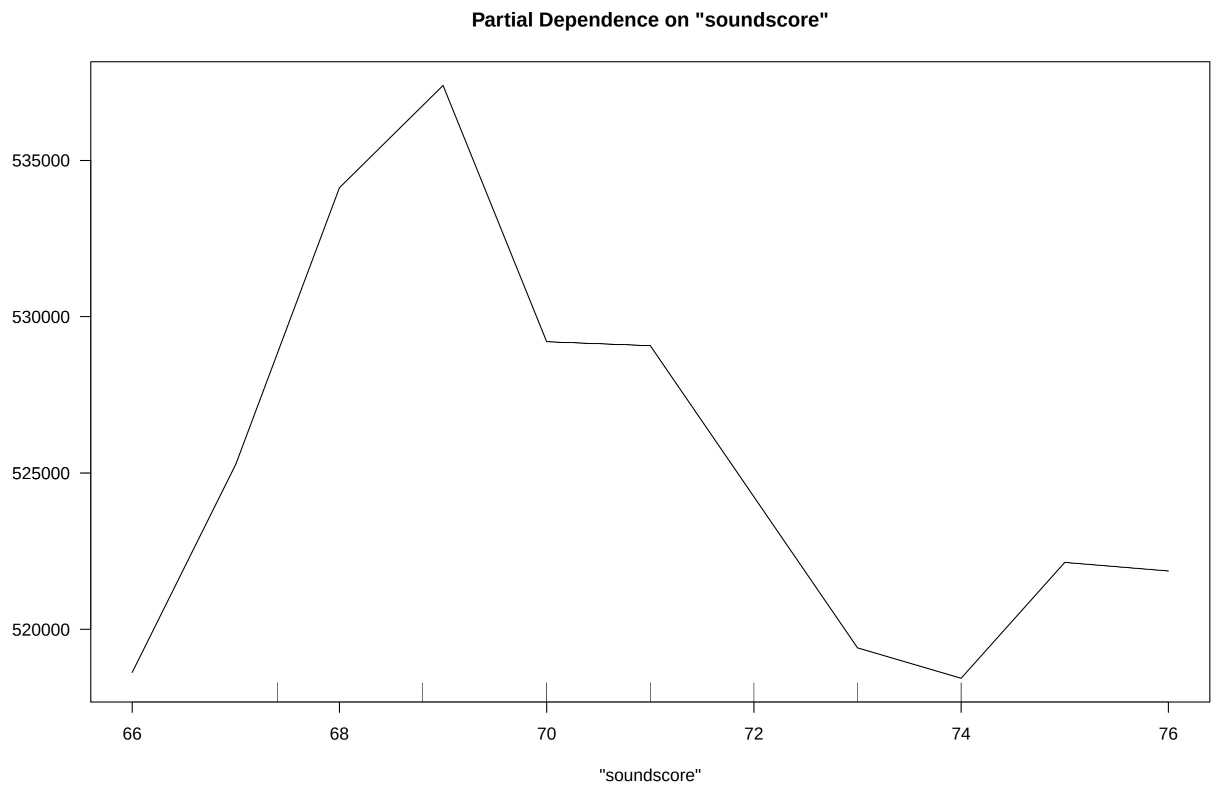
Applying a boxplot to the data refines the picture; the shape of the distribution shows that lowest home values are seen at the loudest and quietest areas.

After the exploratory analysis I looked at a single tree with the CART fitting algorithm and then applied the random forest model.

I then compared the RMSE against the test data. The RMSE for a single tree was 109625 and the RMSE for the random forest was 98982.



Plotting variable importance shows that removing "Soundscore" from the model would increase MSE by approximately 4%.



The partial dependence plot for the random forest model implies the relationship isn't linear. Sale price increases as noise decreases up until around the scores of 68-69 and then decreases.

Conclusion

Noise does appear to have a modest effect on home values, although lower than the other key factors included in the model.

I was surprised to see that the partial dependence plot showed home values decrease at lower noise levels. I'm curious if the rising values at a certain level of noise are partially because homeowners value being close to some of the larger traffic thoroughfares that have more amenities (e.g. the Manor corridor is known for it's concentration of popular bars and restaurants). One area of future investigation would be to enrich the dataset further with local points of interest data and/or neighborhood amenity data.

Appendix

1. Jaffe, Eric. "Why City Noise Is a Serious Health Hazard." Bloomberg.com, Bloomberg, 2015, www.bloomberg.com/news/articles/2015-04-22/why-city-noise-is-a-serious-health-hazard.
2. "Noise management" Gov.uk, 2019, www.gov.uk/government/collections/noise-management