Case Study 3 - Fairness of Assessment Value

Benedikt Farag, Elisa Sommer and Steven Ward
Department of Statistics and Data Science
Yale University

October 7, 2025

1 Objective

A Hamden resident who recently purchased his house had his assessed property value significantly increase within a year of the purchase. He began to question whether the assessed value was fair. He would like to use data to see if there is any evidence he could use to contest his assessed value.

To determine whether the assessed prices are fairly determined, property data for Hamden, CT, is analyzed for properties sold in the recent years. The data was gathered from Hamden online property database. The code for the analysis herein is available in this GitHub repo.

2 Scraping

The assessed values for 2023 and 2024 were available for each of the properties in the Hamden online database. The relevant information was scraped from the .html files that included:

- Total assessed value and associated sub-categories (Building, Extra Features, Outbuildings, Land)
- Most recent `sale_price` and corresponding `sale_date`, i.e. the most recent sale above a threshold of $\$1,000$

A histogram of properties sold (recording only their most recent sale) since 2009 is shown in Figure 1.

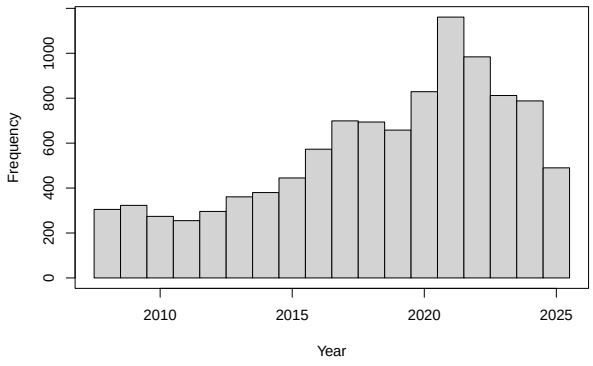


Figure 1: Histogram for most recent sale year from 2009 until 2025*

From Figure 1, there were above 800 sales in the past 5 years (except for 2025 as the data is only until August 13th 2025.

3 Preliminary analysis

The sale price was used as a proxy for the appraised value, and to evaluate the fairness of the assessed value the parameter of interest is

$$\theta = \frac{\text{Assessed Value}}{\text{Sale Price}}$$

and since assessed value is computed as 70% of the appraised value, we can state the null hypothesis as

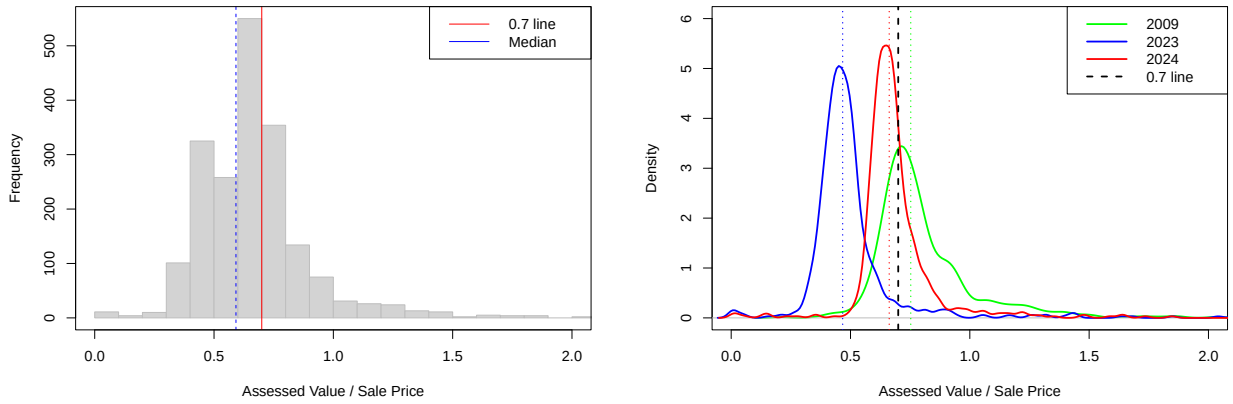
$$H_0 : \quad \theta = 0.7.$$

To ensure an equal comparison, the ratio was computed as

$$\theta = \frac{\text{Assessed 2023}}{\text{Sale Price}} \quad \text{for properties sold in 2023}$$

and similarly for properties sold in 2024. Hence, the properties sold in 2025 were not included in this analysis. The parameter θ ranged from 0 to 5.9 and was heavily right skewed. Some few extreme outliers were removed and the data was filtered for $\theta < 5$.

A comparison of the ratio θ is shown in Figure 2 and Table 1, comparing both the total distribution for both years and separately.



(a) Histogram for the distribution of θ for 2009, 2023 and 2024

(b) Density plots for θ for each year

Figure 2: Overall distribution of θ for both years

Table 1: Summary statistics of θ for 2009, 2023, 2024 and 2023-24 combined

Year	Mean	Median	SD
2009	0.847	0.752	0.388
2023	0.514	0.467	0.289
2024	0.712	0.662	0.300
2023-24	0.613	0.592	0.311

From Figure 2a, the assessed-sale ratio is underestimated with respect to H_0. However, there is a notable difference between the distributions by years as shown in Figure 2b. The 2024 distribution exhibits higher θ values than the 2023 distribution, where the median increases by $\sim 40\%$ between 2023 and 2024. The reference of the 2009 data indicates that the distribution was better centered around the 0.7 target compared to 2023 and 2024, while having higher mean and medians than the target.

The assessed value is very underestimated for 2023 with a median of 0.47 and the distribution shift in 2024 might be due to "correctness" of the assessed value in terms of the market value (appraised value). The median ratio for 2024 is 0.66 that is closer to the desired number 0.7.

So in terms of the objective question, it could be that the Hamden resident bought his house in 2023 prior to the apparent "assessment correctness" indicated in the distribution shift from 2023 to 2024.

4 Hypothesis Tests

To formally test weather the median assessed-sale ratio is different from the desired 0.7, hypothesis tests were performed. Since some of the distributions appear to be approximately normally distributed in Figure 2b, t-tests would be appropriate. To validate the normality assumption, quantile-quantile plots for θ were analyzed in Figure 3

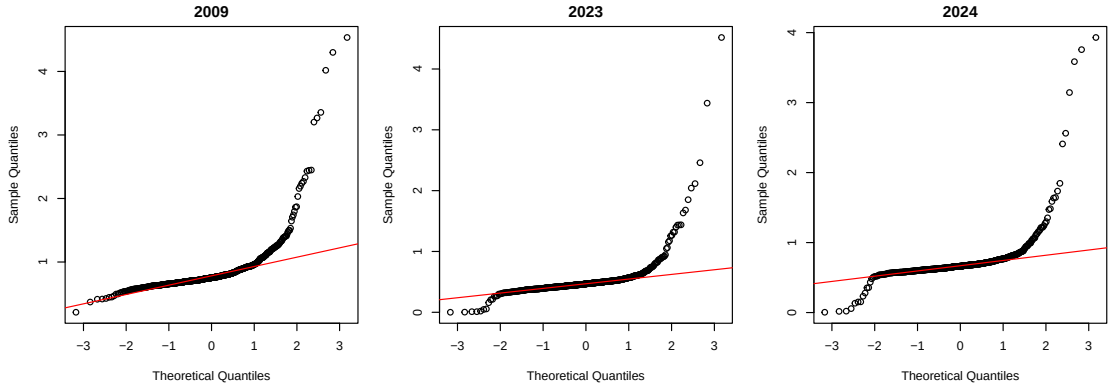


Figure 3: QQ-plots for the θ distribution for each year

The Q-Q plots indicate that the distributions have very heavy right skewed tails. Hence the t-tests are not suitable and nonparametric tests such as Wilcox tests are more appropriate as they do not assume normality.

The Wilcox-tests were all significant at a significance level $\alpha = 0.05$, indicating that the medians were all significantly different from the desired $\theta^* = 0.7$ value. The confidence intervals are plotted for the three years along with the Wilcox median estimate in Figure 4

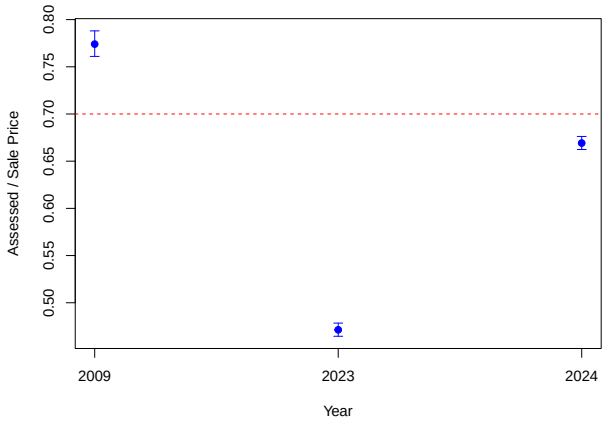


Figure 4: Wilcox-test estimates and confidence intervals

The confidence intervals are very narrow for all the three years, surpassing the 0.7 target in 2009, while being largely underestimated in 2023. The 2024 estimate is closest to the target, while still being significantly different from 0.7. It has to be noted that these three years cannot be interpreted in the same way, as the main focus is the fairness of the assessed value in the recent years in which 2009 is not quite directly relevant. It sheds some light on the distribution in the past, but without years between 2009 and 2023 the interpretation of 2009 is hard.

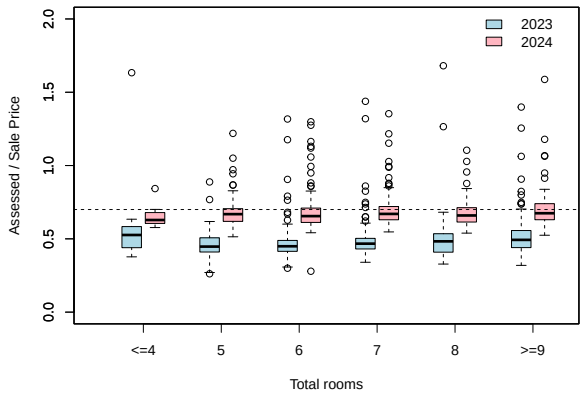
5 Feature Analysis

The assessed-sale ratio $\theta = \frac{\text{assessed value}}{\text{sale price}}$ is now analyzed in terms of few of the relevant property features (total rooms, bathrooms, living area, and sale price). If assessments are fair, θ would be expected to be independent of the different property characteristics, since the sale price is treated as the fair market benchmark. Any systematic trend of θ with property features would indicate that certain types of homes were relatively over- or under-assessed.

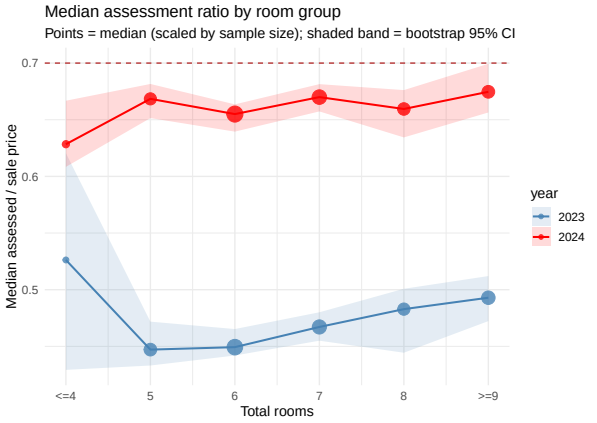
5.1 Total Rooms

We begin by investigating the number of total rooms, with a range between 4 and 9 rooms, as a general measure of property size. Two complementary visualizations are provided for 2023 and 2024

- on the left, boxplots of the ratio θ for each category
- on the right, median values with bootstrapped 95% confidence intervals (CIs)



(a) Boxplots of θ by total room count.



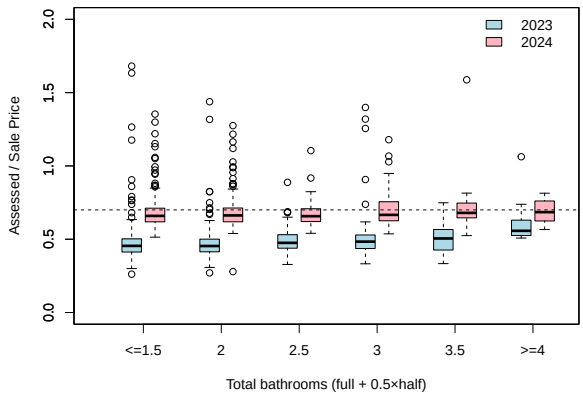
(b) Medians with bootstrapped 95% confidence intervals.

Figure 5: Influence of total room count on θ in 2023 and 2024.

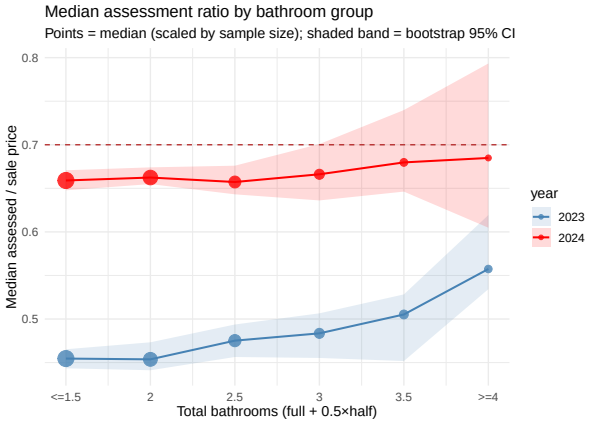
If the assessments were fair, no trend across room counts would be expected. From the boxplot above, the ratios appear roughly similar across categories, but the median plot reveals a mild upward trend in both years, being stronger in 2023. This suggests that homes with more rooms tended to have relatively higher assessed values compared to their sale prices. In 2024, the effect is weaker and the CIs overlap more, indicating a smaller or no systematic difference.

5.2 Bathroom Count

Next, bathrooms were analyzed by combining full and half baths into a single "bathroom number" ranging between 1.5 and 4. Again, both boxplots and median-with-CI plots are shown.



(a) Boxplots of θ by bathroom count.



(b) Medians with bootstrapped 95% confidence intervals.

Figure 6: Influence of bathroom count on θ in 2023 and 2024.

Here the differences between the years are even more pronounced than in Figure 5. In 2023, the median plot shows a clear upward trend: properties with more bathrooms had systematically higher assessed-sale ratios, meaning they were assessed more highly relative to their sale price. The trend is statistically significant, as confidence intervals at the lower and upper ends do not overlap. In

2024, however, the relationship essentially disappears: medians remain nearly flat across categories, with wide overlapping CIs. This indicates that the systematic over-assessment of larger (in terms of bathrooms) properties observed in 2023 was not present in 2024.

5.3 Living Area

Next, the influence of living area on θ was analyzed for the two years. The relationship was found to be nonlinear and was plotted on log-log scale in Figure 7.

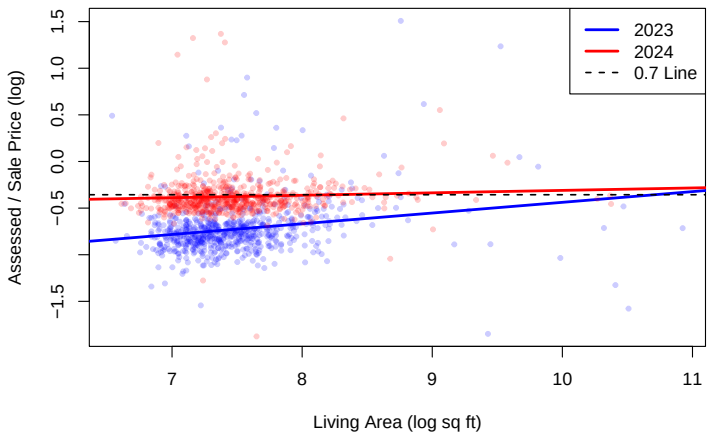


Figure 7: $\log \theta$ versus log-living area

Comparing the two years, there seems to be a notable difference between 2023 and 2024. The slope is significant with p-value of $1.12 \cdot 10^{-7}$ for 2023, while insignificant for 2024 with p-value 0.14 at the significance level $\alpha = 0.05$. Therefore the ratio θ is, on average, not dependent on the living area in the 2024 sales compared to 2023, and there is no apparent effect of living area on the ratio in 2024.

The fitted line for 2024 appears to closely overlap with the 0.7 target line. The strong linear effect with living area in 2023 shows that properties with smaller living area are more underestimated than properties with greater living area. The estimate of the slope for 2023 is $\beta_{2023} = 0.115$, meaning that a 1% increase in living area is associated with an increase of 0.0011 in the ratio θ on average.

5.4 Sale Price

Additionally, the effect of the sale price on the ratio θ was investigated. The relationship was again found to be nonlinear and plotted on log-log scale in Figure 8

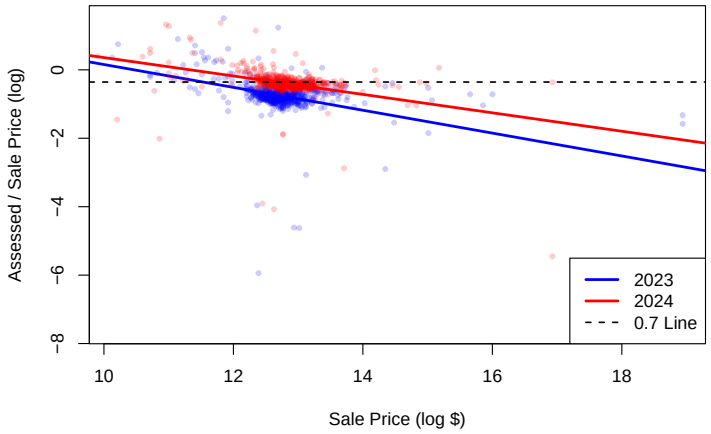


Figure 8: $\log \theta$ versus log-sale price

There appears to be a negative relationship between the log-ratio θ and the log-sale price, for both years. The slopes are relatively similar while the ratio being overall higher in 2024 compared to 2023, as seen before. The dense point cloud for 2024 is centered at the 0.7 target line, but there is still a significant negative trend that might be driven by low- and upper-quantile sale prices. The main majority of the 2023 datapoints underestimate the target 0.7.

Hence, more expensive properties tend to have lower assessed values compared to less expensive properties. This trend is apparent for both years, and is consistent between the years while the ratios generally increased between the years.

5.5 Summary of Feature Analysis

5.5.1 Size-related effects (rooms, bathrooms, area)

Under fair valuation, the assessed-sale ratio $\theta = \frac{\text{assessed}}{\text{sale price}}$ would be expected to be approximately constant across property characteristics. Across all three size measures we examined (total rooms, bathrooms and living area) the same qualitative pattern was observed: θ increases with size. This trend is present in both years but is notably stronger and statistically significant in 2023 (as seen in the median-with-CI plots), while in 2024 it is weaker and often not distinguishable from flat. In other words, larger homes tend to have higher multiples (higher assessed value relative to sale price), especially in 2023.

5.5.2 Effect of Sale-price

At first glance, the *negative* relationship between θ and the *sale price* seems contradictory on the log–log scale.

One possible explanation is that the state's assessments are partly based on a regression model that predicts sale prices from property features such as area, bathrooms, and total rooms. For the homes that did actually sell, the fitted value from this model may then have been linearly combined with the actual sale price, before applying the 70% factor, i.e. the assessed price was determined on a mix of fitted value of the regression and actual sale price. If this is the case, it would explain why θ tends to increase with size-related features (since the model strongly uses them), while at

the same time showing a downward trend with respect to the actual sale price (since for more expensive homes the fitted value may not keep up with the observed price). In 2024, the patterns are weaker, suggesting that the assessments were adjusted more closely to actual sale prices.

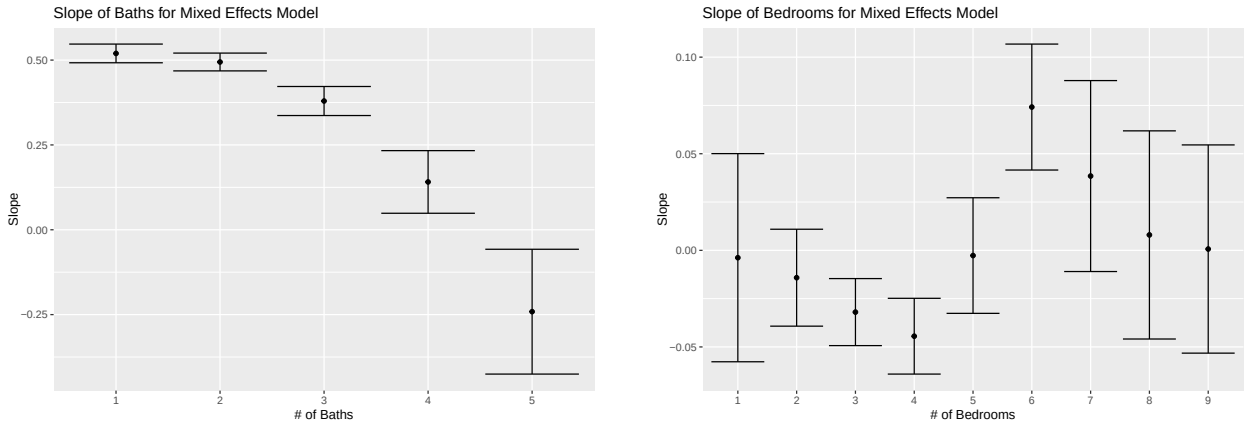
Other possible explanations:

- **Different homes sold:** The mix of homes sold in each year can still matter, even though the sample size is large (over 800 sales in both 2023 and 2024). If relatively more large homes happened to sell in 2023, that could reinforce the upward trend with size in that year.
- **Regression to the mean:** Very low sale prices might not always reflect true market value (for example, discounted sales to friends or family, or distressed transactions). In those cases, the assessed value would not be adjusted downward as much, leading to higher multiples at the low end. This could help explain why we see higher ratios for cheaper properties, consistent with the downward slope versus sale price.

6 Regression Analysis

To further analyze the effects of bathrooms and bedrooms on the parameter θ, the mixed effects were analyzed. Mixed models do not rely on independence of observations are are thereby very useful to analyze the difference between years.

Thus, a mixed effects model on Assessed Total was fitted to the data. This model used the year, living area, acres of land, number of bedrooms, and the number of bathrooms as predictors of the log of the assessed total, with random effects for each year (2023 and 2024) on number of bedrooms and the number of bathrooms. Living area and Land area were modeled as fixed effects. These random and fixed effects were chosen to see if there was a change in how the Hamden assessment model performed from 2023 to 2024. Specifically, checking to see if the number of bedrooms and the number of bathrooms had any change on the assessed values.



(a) Slope with standard error for number of bathrooms in 2024.

(b) Slope with standard error for number of bedrooms in 2024.

Figure 9: Slopes for number of each type of room in 2024 Assessed Value Model

Figure 9 shows the slopes for each number of each type of room in 2024 as compared to 2023. Therefore, any non-zero slopes indicate a change from 2023 to 2024. Comparing the two plots, the

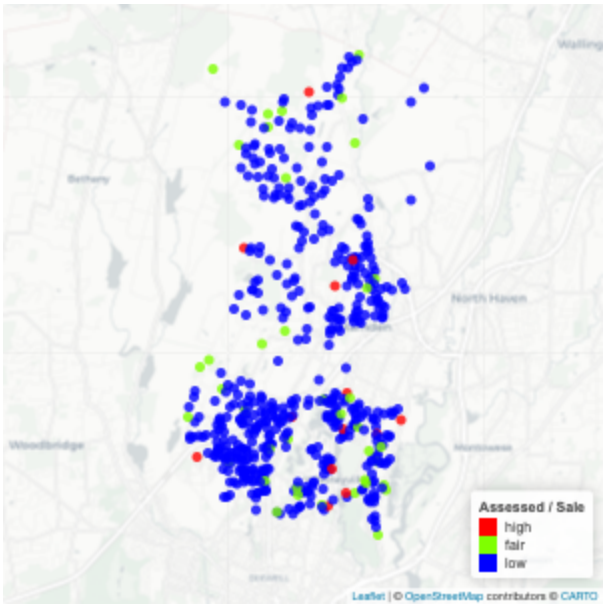
change in the bathroom slope is more prominent than the bedroom slope. While the estimates of the bedroom slopes are all non-zero, 6 out of the 9 of the groups include zero with their standard error ranges. Because of this, one cannot assume that all these slopes are significantly different than 0. This means the number of bedrooms is accounted for the same way in the 2024 model as it is in the 2023 model. In other words, the bedroom slope does not change much, if at all, from 2023 to 2024. However, all the bathroom categories exclude 0 in their standard error ranges. Because of this, the Hamden Assessment model appears to treat the number of bathrooms differently in 2024 than in 2023. In other words, the number of bathrooms seems to influence the change in the assessed values from 2023 to 2024. This impact would also be a positive change since the slopes are positive. This aligns with how the assessed values shifted from 2023 to 2024, in Figure 2b.

7 Spatial Analysis

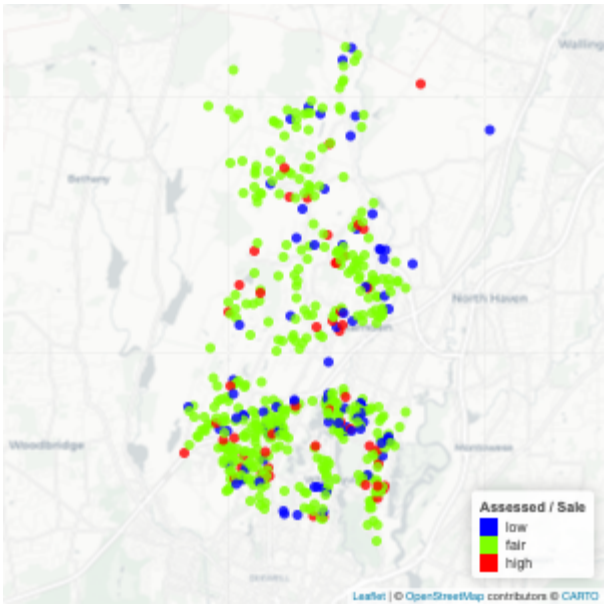
The distribution of the assessed-sale ratios θ were also viewed geographically. The spatial analysis of the assessed values did not yield any telling results. When comparing the 2023 and 2024 plots, no spatial conclusion can be easily drawn. There are no clear areas that have higher or lower θ The only obvious observation from the plots in Figure 10 is that the assessed-sale ratio is lower across the board in 2023 than in 2024.

The thresholds for the low-fair-high categories were selected from thresholds as:

low	**fair**	**high**
$\theta < 0.6$	$0.6 \leq \theta \leq 0.8$	$\theta > 0.8$



(a) Assessed-sale ratio θ in 2023.



(b) Assessed-sale ratio θ in 2024.

Figure 10: Spatial distribution of the assessed-sale ratios for 2023 and 2024

Again, the plots clearly show the general trend that in 2024, most of the Assessed values were fair (between 0.6 and 0.8) and in 2023 most of the Assessed values were low. There were no specific streets that had consistent ratios category, no neighborhoods, and no general areas. It truly seems random. The complete increase in the ratio θ between 2023 and 2024 likely stems from a change in Hamden's approach on how the appraised values are calculated.

8 Summary

The goal of this project was to analyze the "fairness" of the assessed values of properties in Hamden, CT. The reason the assessed values were under scrutiny was due to a homeowner saw a large jump in the assessed value of their home just one year after purchasing. The analysis conducted herein has revealed key characteristics in the fairness of the assessed values. There was a clear distribution shift to greater assessed values of the Hamden properties in 2024 compared to 2023. This shift seemed to be stemming from a "corrected" assessment that were greatly underestimated in 2023. The 2009 data was used as a reference for the past evaluations, that was however not useful on determining the fairness of the current assessment model by Hamden.

This conclusion was shown in many ways, starting with Section 3 and the preliminary analysis. The simplest metric for the fairness of assessed values is the summary Table 1. The mean and median in 2023 are well below 0.7 (the true hypothetical value). However, the mean and median for 2024 fall above and below the target, respectively, and has a distribution better centered around the 0.7 target.

The next way this conclusion came to be was in Section 4, where hypothesis test were performed for the assessed-sale price θ. The distribution is highly right skewed and the normality assumption thereby fails. Therefore, a Wilcox test was conduced, the median assessed-sale parameter for each year were all determined to be significantly different the 0.7. However, looking to Figure 4, 2024 is much closer to the target value than in 2023. So, while neiter year may be exactly fair, 2024 made a step in the right direction compared to 2023.

In Section 5, specific features of the homes were tested to find the cause of this upward shift in assessed values. Starting with total rooms, there is a stronger trend in the number of rooms in 2023 than in 2024. This again indicates that 2024 is more fair than 2023. For number of Bathrooms, this same trend was found again, i.e. that there is a stronger trend in 2023 than 2024 indicating again 2024 is more fair. Both trends in 2023 tended to have higher ratios for higher numbers of rooms and bathrooms.

In section 6, a mixed effects model was fitted on the log of assessed total. The random effects were number of bedrooms and number of bathrooms. They were random for each year (2023 and 2024). The fixed effects were living area and land area. The random effect slope for number of bedrooms did not seem to change from year to year, but the slope for number of bathrooms did seem to change in a positive way. This could have been part of the update in the Hamden assessed values model. Since these slopes were mostly positive, it indicates that the change in the slope for the number of bathrooms could push up the assessed values for homes in Hamden.

Lastly, a spatial analysis was performed. While no insights were found to which areas yield higher or lower assessed values, it is clear that most of the houses sold in 2023 had low assessed values whereas most of the houses sold in 2024 had fair assessed values. This supports the conclusion that in 2023 the assessed values were low and they jumped up to be more fair in 2024.

In conclusion, there was some unfairness identified in the 2023 data that decreased or even totally vanished in 2024. This was revealed from the analysis for rooms, bathrooms and living area. However, there is a strong negative linear trend between the $\log \theta$ and log-sale price. This shows that the assessed value is generally lower for more expensive properties, that might be "unfair" for less expensive properties, since the taxes are directly computed as a fraction of the assessed values. Hence, the more expensive properties seem to be under-taxed due to the under-estimated assessed value.