

Analysis of Residential Real Estate Data From Connecticut*

Siling Guo, Megan Joseph, John Sawicz

November 16, 2025

1 Introduction

Real estate trends change over the years. The most notable changes have been during the 2008 economic crisis and COVID-19. We decided to investigate real estate in Connecticut, USA to see how it's changed over time and how it differs across cities.

We first clean the data and calculate descriptive statistics. Then, we make graphs of the categorical variables and continuous variables based on the cities we picked. In our advanced analysis, we make heat maps showing the densities of sale price and a multiple linear regression model.

2 Data

The data we used for this analysis is Real Estate Sales data from 2001-2023 from the State of Connecticut's Office of Policy and Management. The sale price of each property is at least \$2,000. Each row is a property which contains information of the town, address, date sold, property type (residential, apartment, commercial, industrial or vacant land), sale price, assessed value, and latitude and longitude coordinates. For the purposes of this analysis, we mainly focus on residential properties and the columns town, property type, sale price, assessed value, and coordinates. Additionally, we picked four cities to focus on: Stamford, Westport, Cheshire, and Sprague. This was done because of the large number of data points and to investigate any differences in towns with varying levels of median income. Westport, CT has the highest median income at \$250,001, then we picked Cheshire, CT at \$150,787 for upper middle, Stamford, CT for lower middle, and Sprague, CT for the lowest. We also wanted to sample towns with different populations and densities. (2025, n.d.).

*Project repository available at: https://github.com/meganajoseph/167r_project.

2.1 Data Cleaning

Many of the data points were missing or were empty characters, so we dropped those rows. Since our goal is centered on residential properties, we filtered out rows that were not residential. Additionally, the Sales.Ratio column needed to be transformed into a numeric value.

2.2 Descriptive Statistics

We analyzed the mean, minimum value, maximum value, first quantile, median, and third quantile of the Sale.Amount, Assessed.Value, and Sales.Ratio columns. The information is summarized in the table below.

Column	Mean	Minimum	Maximum	1st Quartile	Median	3rd Quartile
Sale Amount	7.7×10^5	2160	7.2×10^7	3.1×10^5	5.04×10^5	8.2×10^5
Assessed Value	4.6×10^5	0	2.5×10^7	2×10^5	3.3×10^5	5.3×10^5
Sales Ratio	0.8	0	291.94	0.55	0.64	0.76

3 Graphs

3.1 Categorical

We created bar plots to see the number of entries per town of interest and per residential type.

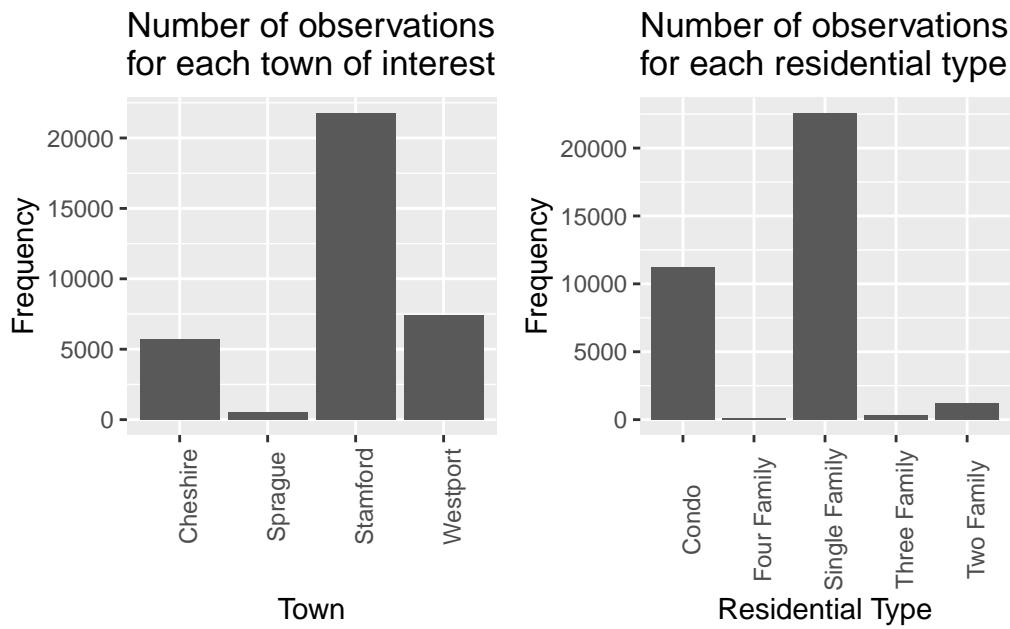


Figure 1: The plot of the left shows that the number of observations is greatest for Stamford and lowest for Sprague. The plot of the right shows that There are more entires for Single Family and Condo properties and very few for Two Family, Three Family, and Four Family.

Since Stamford is a big city with a large population, it makes sense for it to have the most entries. On the other hand, Sprague is the opposite as a small town with a small population which accounts for the low amount of sales. Most housing are Single Family or Condos. It is rare to see Two Family and above sized homes being built.

3.2 Continuous

3.2.1 Distribution of Sale Price



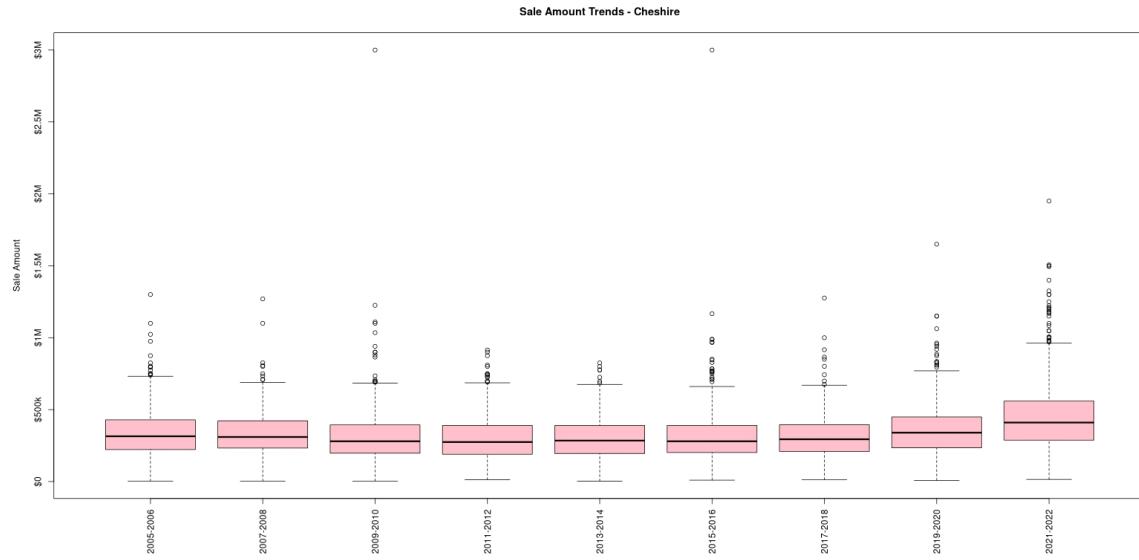
Figure 2: Sale price seems to be centered at around $\$e^{13}$.

We took the log of the sale price because it is extremely right skewed otherwise. We see that the data is centered at around $\$e^{13} \approx \$442,413$.

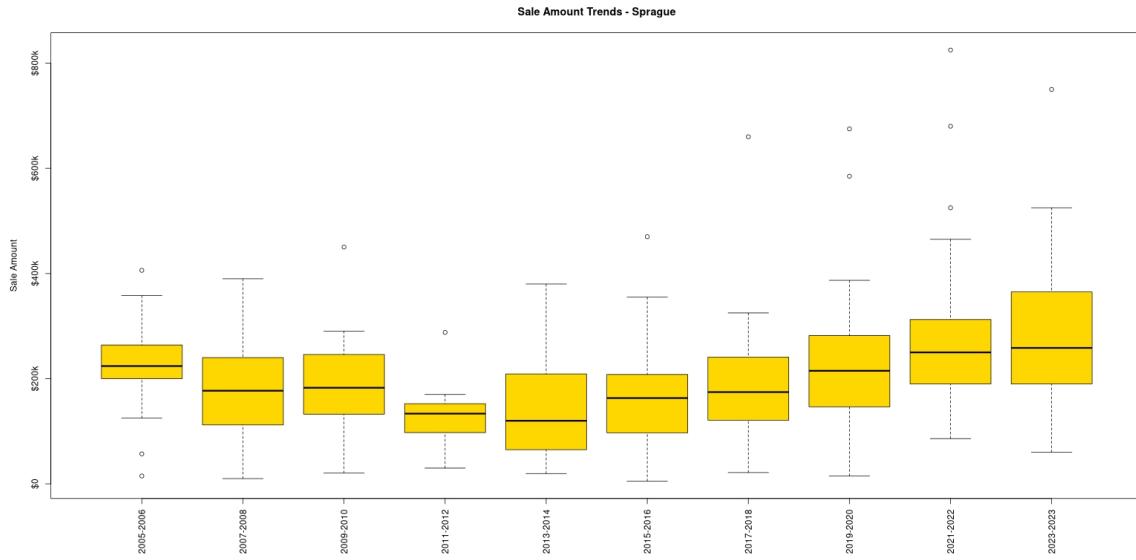
3.2.2 Box Plots By Town

We constructed a series of boxplots and scatterplots for each of our continuous variables (sale price, assessed value, and sales ratio), and then repeating this step for each of the four towns selected (Stamford, Westport, Cheshire, and Sprague). We decided that aligning biannual boxplots vertically and then sorting them in time order would demonstrate trends over time in a more effective manner.

Sales Amounts:



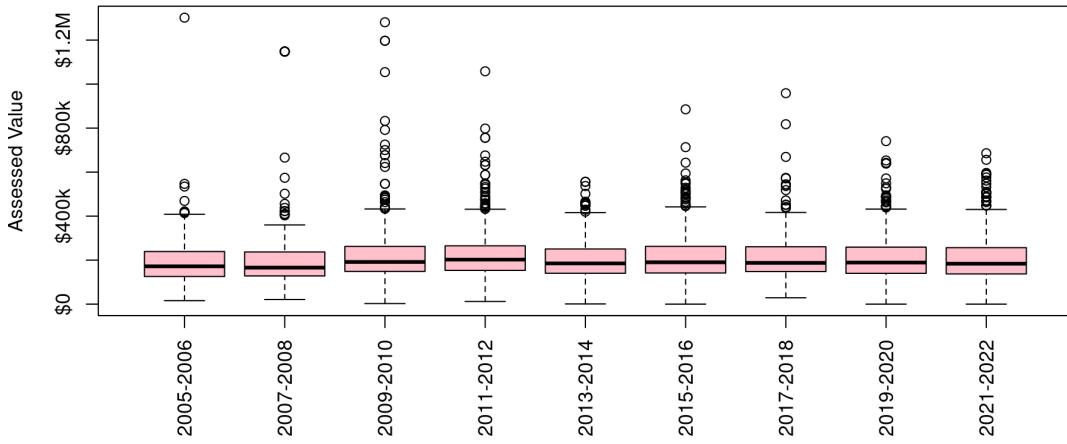
Prices in Cheshire seemed to slump for longer after 2008, both in the outlier spread and the main quartiles, not really rebounding until the Covid-19 pandemic. It is farther away from major metropolitan areas and the coast, with a smaller population. It appeals to buyers with more modest means.



Sprague, the smallest of the towns we chose, displayed more year to year variability and a great deal more market driven variability than larger towns. Part of that is likely due to lower sales volume, and perhaps since it is less prosperous it was a relative bargain and accessible to more buyers.

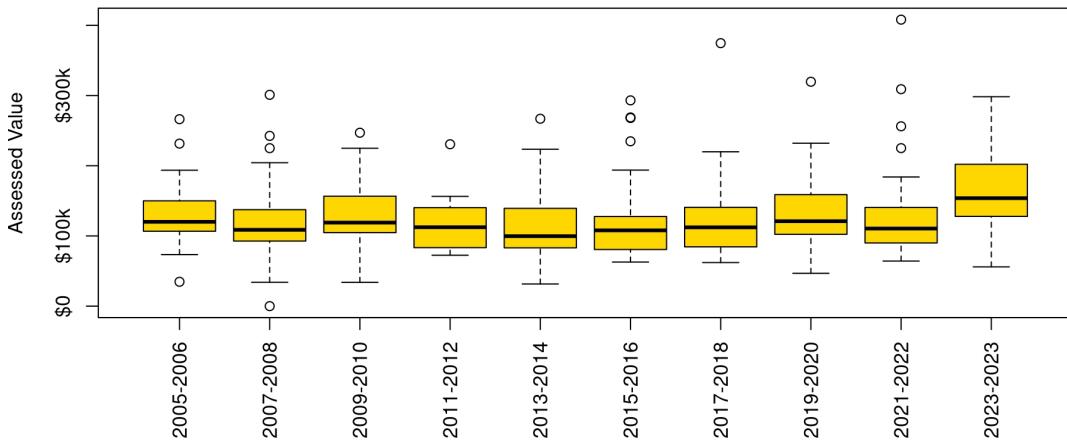
Assessed Values:

Assessed Amount Trends - Cheshire



The assessments tend to display less year on year variability than the actual sales prices. The IQR remain remarkably constant, with the most visible variability in the outlier field. This suggests that for all but the most expensive homes, the assessor is simply following a universal formula to calculate the value of a house. With the most expensive, it could be a different metric that takes in unique features that only wealthier people have, or it could be that fewer of these sales cases more variation by year.

Assessed Amount Trends - Sprague



Sprague is an interesting case, because its a relatively poor area with fewer residents. The high assessment variability here can probably be chalked up to fewer units assessed every year. There could also be the opposite issue to the high outliers. The low outliers may be in a uniquely bad condition, each requiring its own consideration.

Sales Ratio:

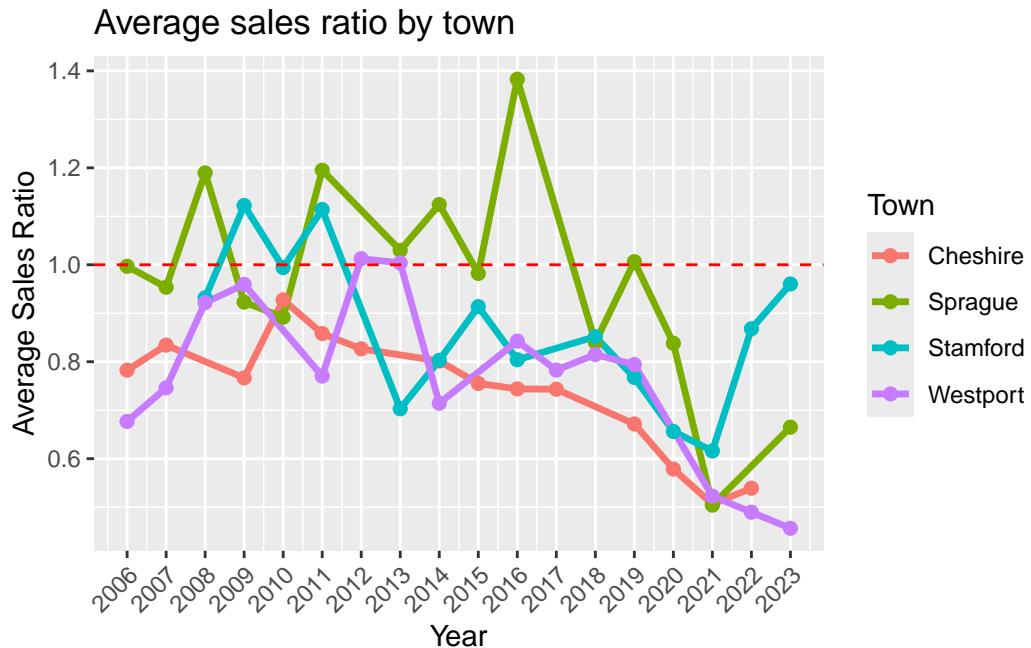


Figure 3: Average sales ratio by town:Average Sales Ratio (ASR) = Average Assessed Value (AAV) / Average Sales Amount (ASA). ASR > 1 means the Average Assessed Value is higher than the Average Sales Amount ; ASR = 1 means the AAV is equal to the ASA; ASR < 1 means the AAV is lower than the ASA.

?@fig-sr-plot Westport is a wealthy town with high property values, and Cheshire is a upper-middle-class town. Stamford is a urban city where near NYC and its economy is closely related to NYC. Sprague is a small town with lower-middle-income. In general, The well-off town (Westport/Cheshire/Stamford) experience smaller sales ratios. during the 2008 financial crisis, the less economically developed town (Hartford) and the NYC-dependent-economic-structure town (Stamford) experience greater market volatility. Considering the time-lag factor, During the 2007–2008 subprime mortgage crisis, housing prices plummeted, causing the ASR in many towns to rise rapidly, even faster than the speed of assessment adjustments. After 2009, the ASR in most towns declined because the real estate market began to recover, and housing prices increased faster than the assessment values. During the global pandemic in 2020, the ASR of all four towns showed a significant downward trend, reflecting a surge in housing

prices which causing by the extremely low mortgage interest rates and the increase in housing demand driven by the work-from-home trend.

4 Advanced Analysis

We decided to make a heat map of sale price, assessed value, and sale ratio for the entirety of Connecticut and the cities we selected.

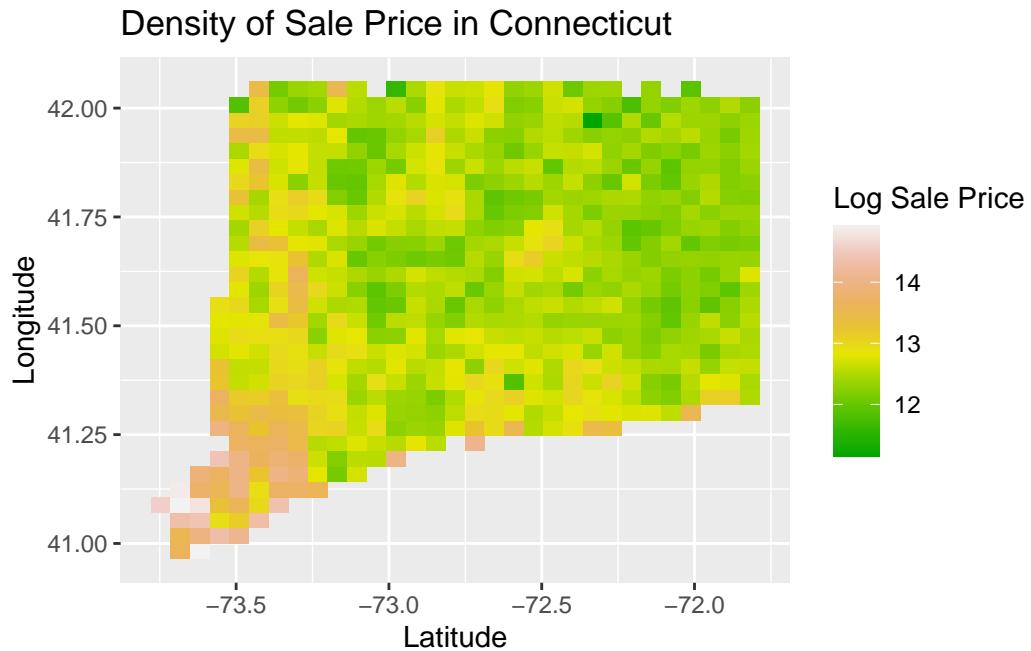


Figure 4: Heat map showing the densities of sale price throughout Connecticut, USA

Figure 4 shows the densities of sale price for Connecticut. For the majority of Connecticut, the sale price is uniform. However, the closer you get to New York, the higher the sale price is.

Figure 5 shows the densities of sale price in the four cities we picked. Stamford and Westport seem to have the most data points, with Sprague being the most sparse. Sprague, Westport, and Stamford are more uniform whereas Cheshire has more variability in prices.

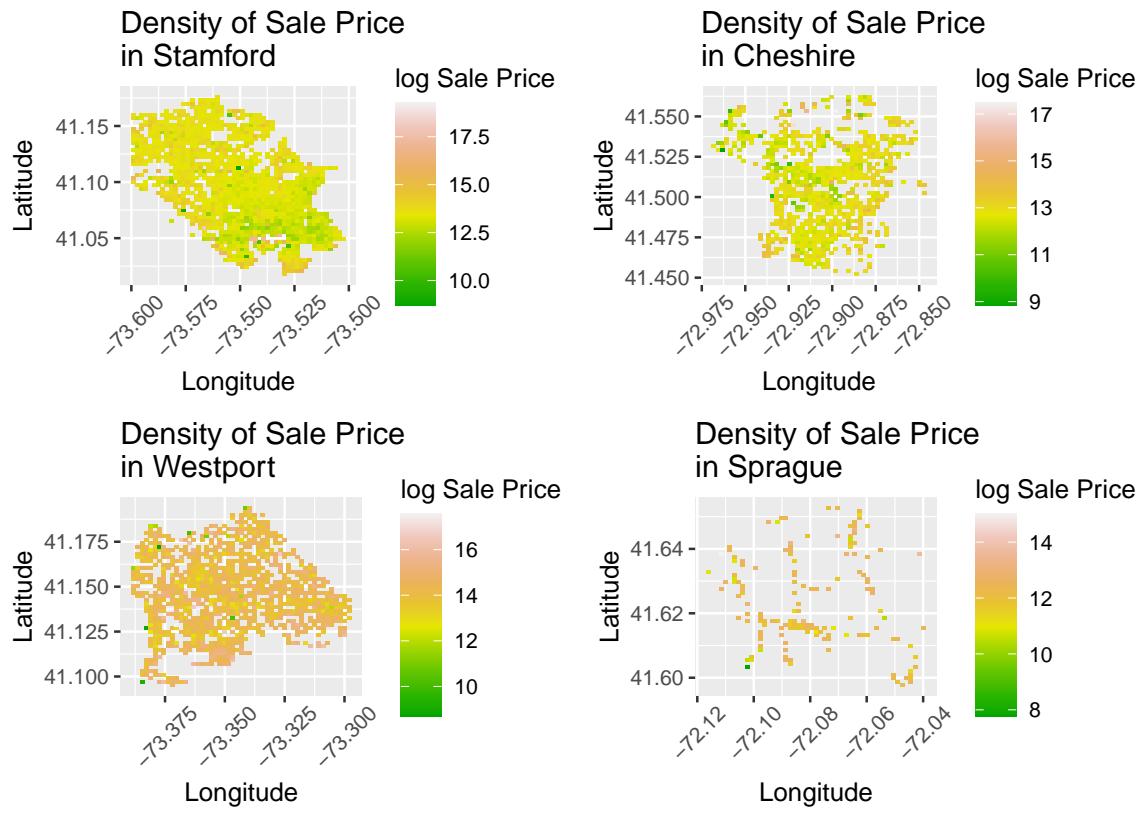


Figure 5: Heat maps of sale price for each city of choice: Stamford, CT on the top left, Cheshire, CT on the top right, Westport, CT on the bottom left, and Sprague, CT on the bottom right.

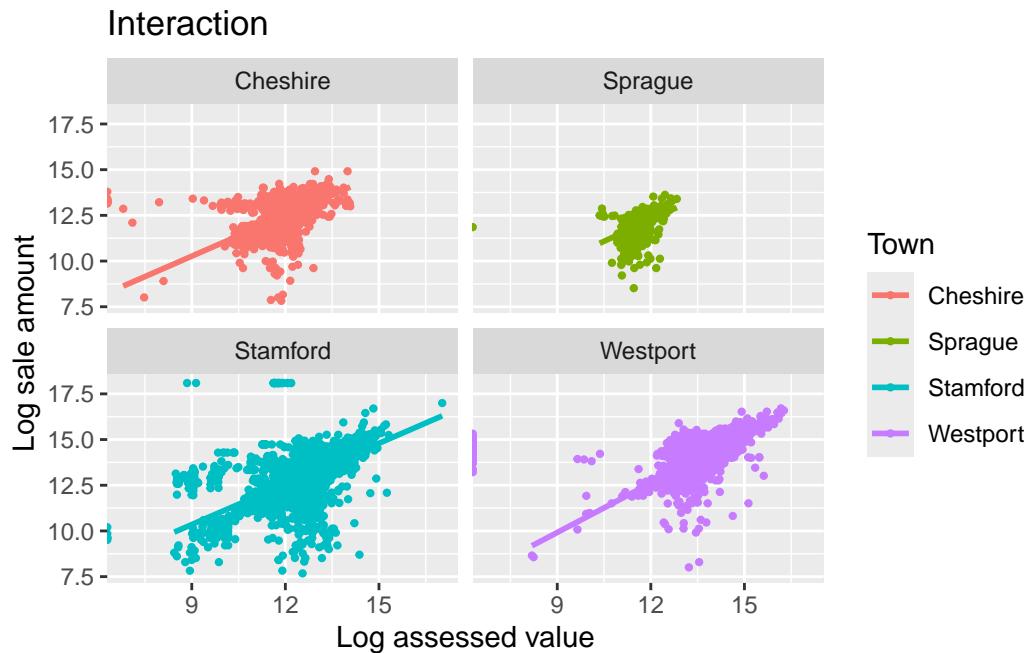


Figure 6: Scatter plot of regression model

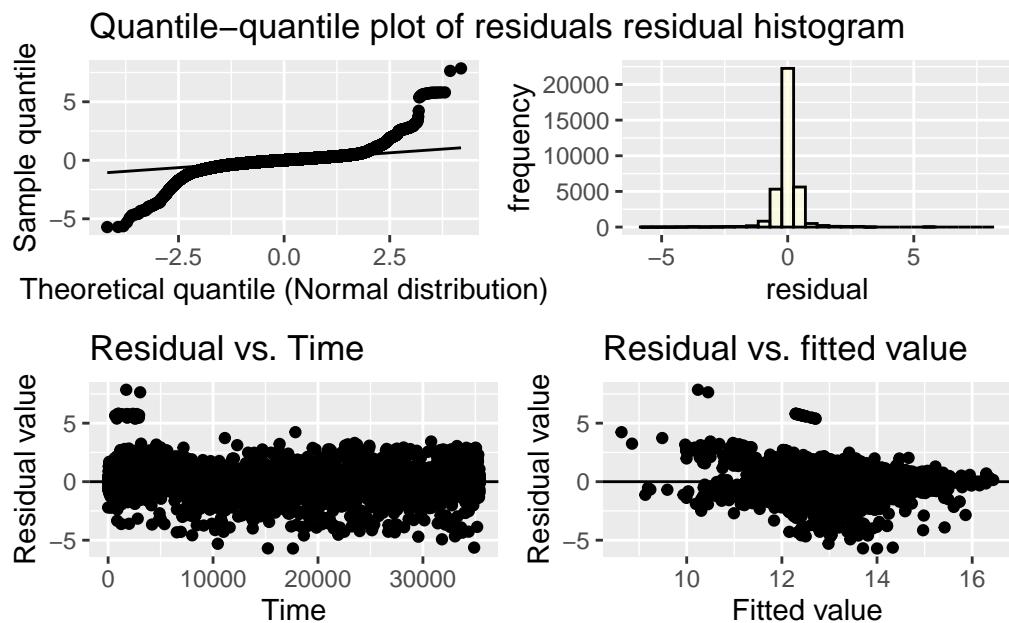


Figure 7: Assumptions

5 Conclusion

References

2025, Data Commons. n.d. “Place Rankings - Data Commons.” *Data Commons*.
https://datacommons.org/ranking/Median_Income_Household/CensusCountyDivision/geoId/09?h=geoId%2F0915021860&unit=%24.