ECNS 460
Abe Seybert
Ethan Carney

Our topic involves using spatial analysis to analyze real estate prices across property zones in Bridgeport, CT. Our motivation is to use raster data to visualize geo-spatial differences in property values and their relation to zone classifications designated by the city. This is a descriptive analysis, as we aim to quantify the relationships between variables across space and time.

Our datasets came from two sources: A record of real estate sales in Connecticut that spans from 2001-2022 (from the official CT.gov website), and a GIS zoning data set with zone classifications and locations of zones in Bridgeport, CT (from the City of Bridgeport website). The reason for selecting Bridgeport is explained below. These are the two websites:

https://data.ct.gov/Housing-and-Development/Real-Estate-Sales-2001-2022-GL/5mzw-sjtu/about_data

https://city-of-bridgeport-gis-hub-bridgeportct.hub.arcgis.com/datasets/d04302caa02146cb9e41af681a23bdf0/about

In this data cleaning process, we began by loading the housing sales dataset that contained records of housing sales in Connecticut from 2001-2022. Given the large size of this dataset (1 million+ observations), it seemed best to focus on one city, so we landed on Bridgeport, as it had the highest number of observations of any town. We then imported the GIS zoning data and visualized them using the viewer in R to better understand the spatial distribution of zones. The property sales data set had the following variables: house serial number, listing year, date the sale was recorded, town, address, assessed value, sale amount, sale ratio (Ratio of the property sales price to its assessed value), property type, residential type, non use code, assessor remarks, OPM remarks, and location (longitude and latitude points). There were a large portion of observations that had addresses for the property, but no location point. To address missing geo-spatial locations, we used the Google Geo-Spatial API to geocode addresses that didn't already have coordinates, implementing error handling for cases with incomplete or ambiguous addresses. By using parallel processing, we were able to handle the geocoding efficiently across a large number of records (although it took a while still).

The zoning data had the following variables: Name of zone (city designated code), Zone classification, and geometry (multipolygon of the zone's geographic location and shape).

This data set needed little cleaning, although we removed the altitude portion of the geometry to work with simply longitude and latitude values.

We then performed a spatial join of each property sale with the zone it was located in, and preserved both the property location and the geometry of the zone it is located in. After combining the property data with the zoninging data, we refined the dataset further by removing unnecessary columns, such as comments and temporary variables used for geocoding. We organized the data so that each observation represented a unique property sale, and created a primary key based on the property's serial number and sale date. For consistency, we converted categorical variables to factors and truncated descriptions in the `Non.Use.Code` field, removing any leading zeros where necessary. We also set missing values to `NA`, replacing zero values in the assessed and sale amounts with `NA` to indicate cases where no assessment or sale occurred. Finally, we saved the cleaned dataset for further analysis and created a visualization showing property counts by zoning classification, which will help us gain insights into spatial and economic patterns across different zones in Bridgeport.

Figure 1

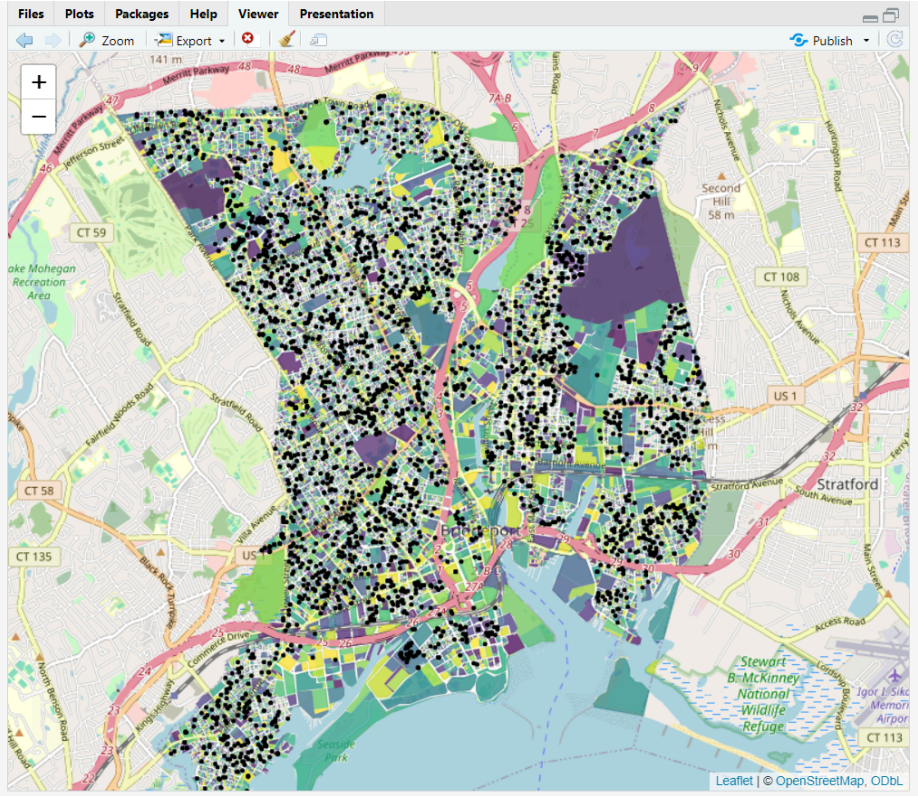

Figure 1 is a screenshot of an interactive map that overlays houses sold on the different zones in the data set. This map offers unique pricing insight on all different zones in bridgeport. This was created using the Google Geo-Spatial API to geocode addresses mentioned above.

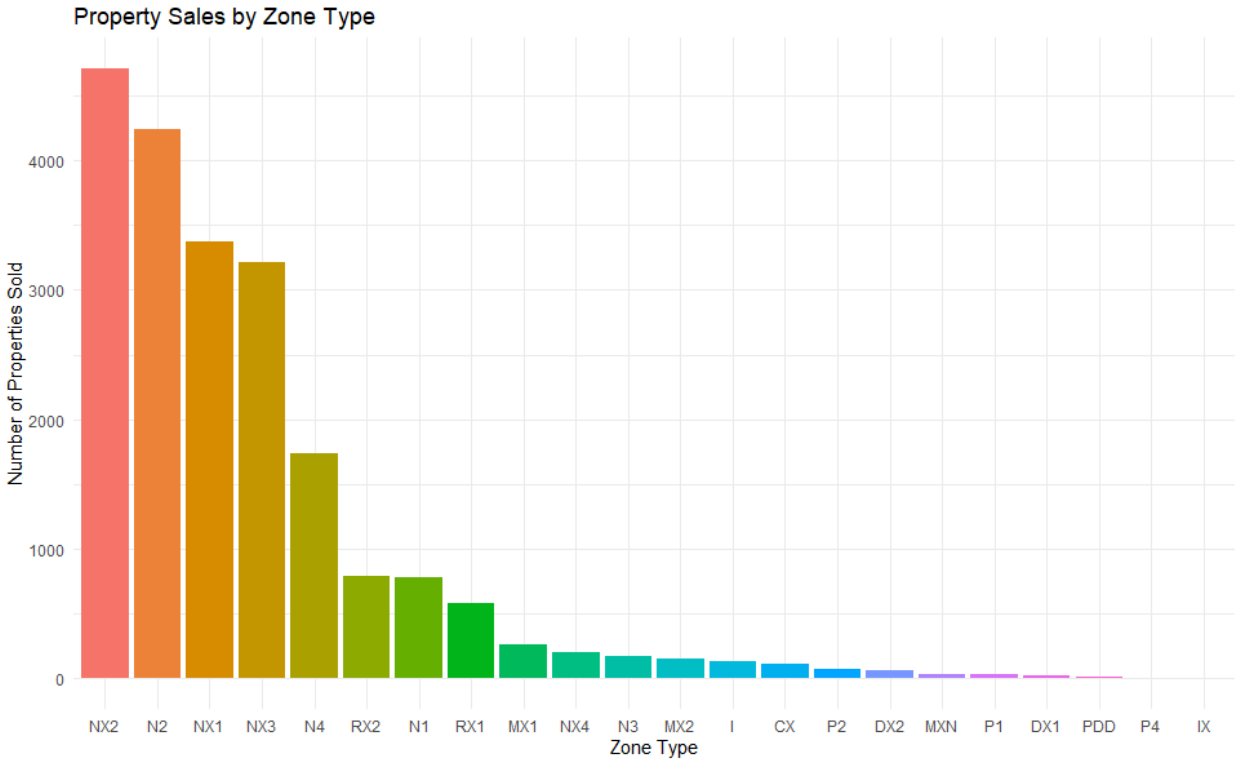

Figure 2

Figure 2 shown above illustrates the quantity of houses sold in each zone in Bridgeport, arranged in decreasing order to highlight zones with the highest sales volumes. This data helps us identify which zones are experiencing the most market activity, offering insights into areas with potentially high demand or turnover.

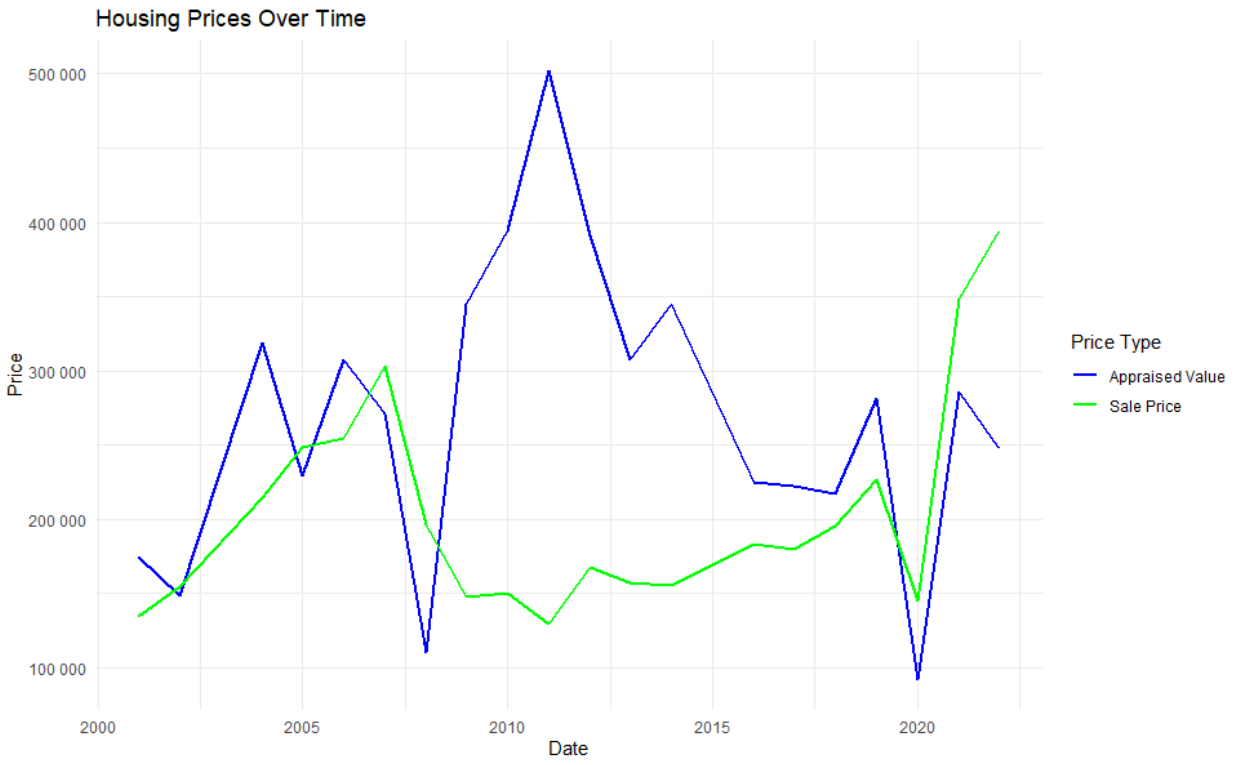

Figure 3

Figure 3 presents a time series comparing appraised values and actual sales prices over time. The two metrics generally track closely, reflecting alignment between property appraisals and market sales trends. However, a noticeable spike in appraised values around 2008 due to the financial crisis. We also can see a notable dip in appraised price as well as selling price around 2020, presumable due to Covid-19.

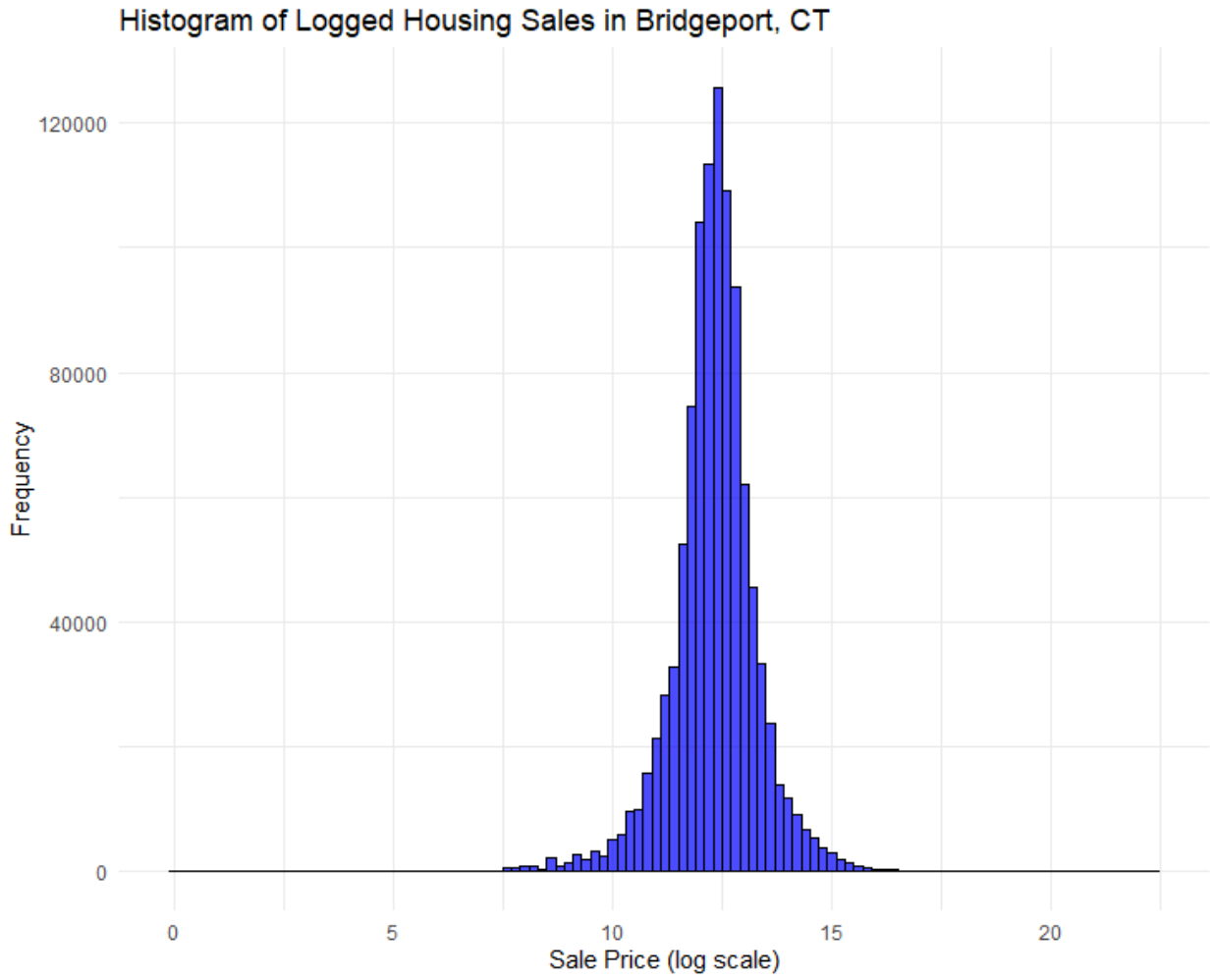

Figure 4

Figure 4 illustrates the distribution of loggrf housing sales in Bridgeport, CT. The original data showed a high skew due to a few extreme sale prices, which obscured the overall pattern. To address this, we applied a log transformation, resulting in a more interpretable histogram. The transformed histogram now displays a relatively normal distribution, reflecting a more typical range of housing sales prices

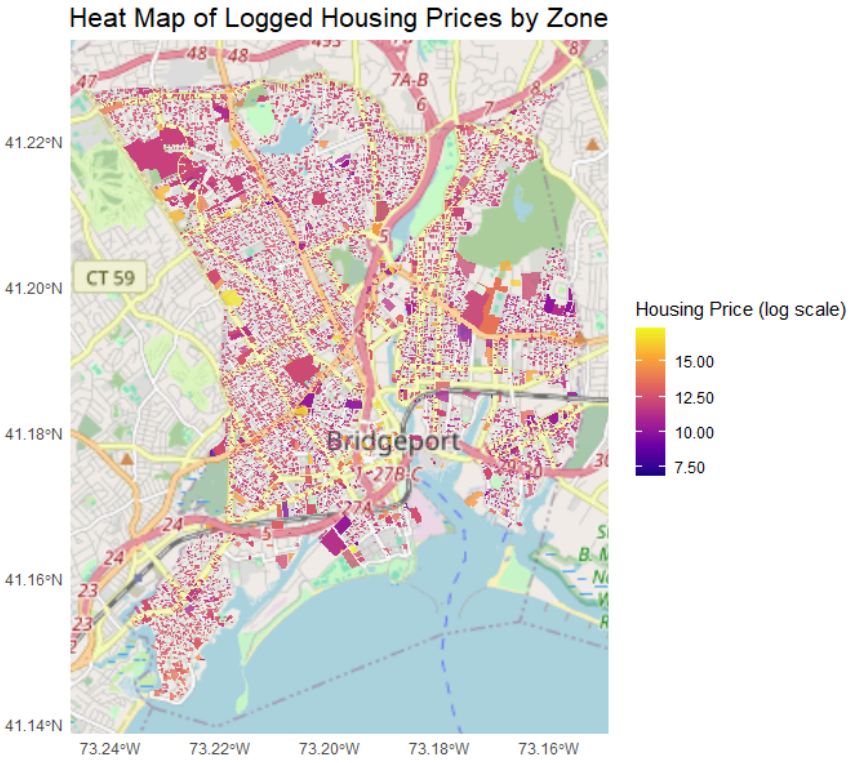

Figure 5

Figure 5 displays a heatmap of the Bridgeport area, with color variations representing the distribution of house prices across different city zones. Warmer colors (e.g., reds and oranges) indicate higher-priced sales in zones, while cooler colors (e.g., blues and greens) represent lower-priced sales in zones. This visualization allows us to observe patterns in housing prices geographically, providing insights into how property values vary by zone.

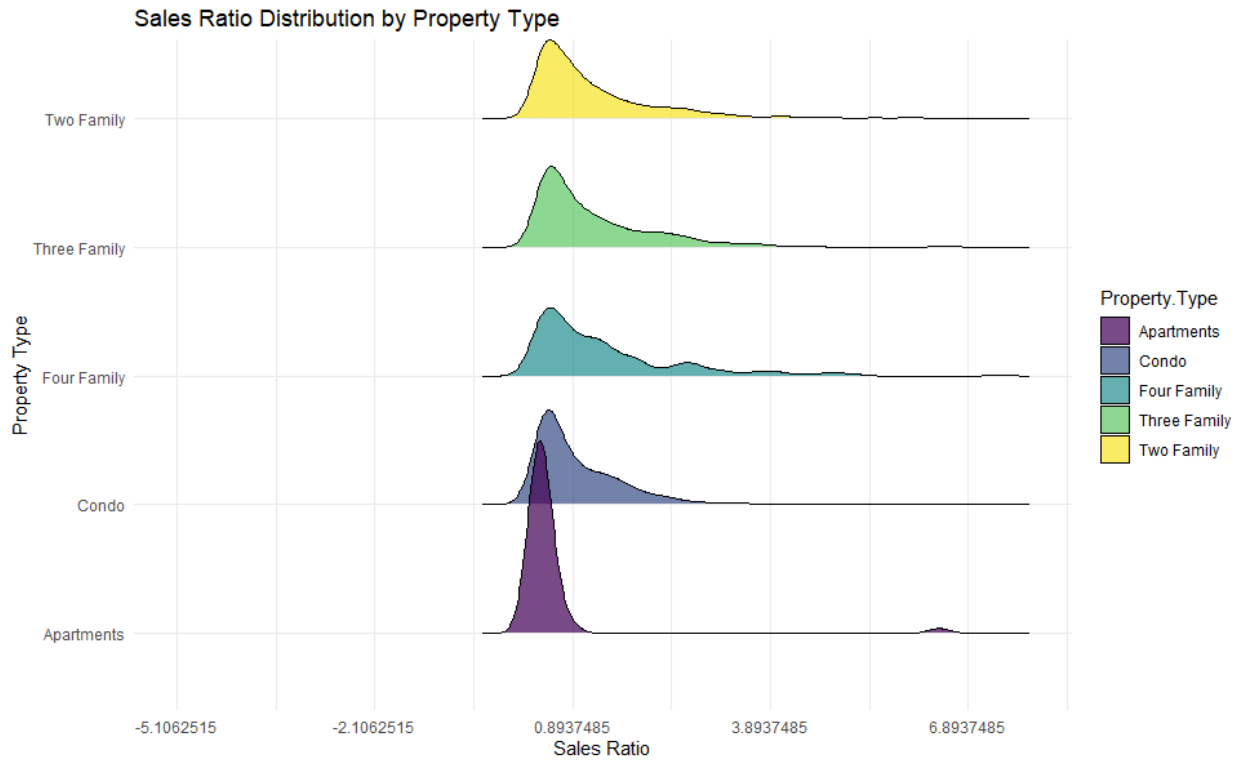

Figure 6 presents the sales ratio by property type, revealing trends across different property classifications. The plot shows similarities among Two Family, Three Family, and Four Family properties, as well as Condos, which all display a broad range of sales ratios. However, Apartments exhibit a tighter cluster around a sales ratio of 0.8, suggesting that most apartment properties in Bridgeport tend to sell for prices close to a consistent benchmark relative to their assessed values. We can also see that there is an apparent right skew in all of the categories shown above. This insight can help identify property types with more stable market performance and assist in appraising and pricing strategies.