

30538 Final Project:

Lauren Laine, Ella Montgomery, and Griffin Sharps

2024-12-06

Github names: laurenaine, montgomeryel, and Griffin Sharps. All Members are in Professor Shi's T/TH 3:30PM section.

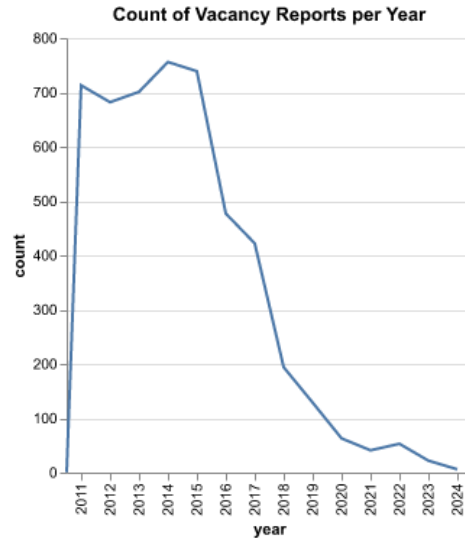
For our final project we were interested in exploring how the prevalence of vacant buildings does (or does not) relate to housing prices and income in Chicago through spatial analysis. More specifically, our research question was “how can we visually map the relationship between housing prices, income brackets, and the number of vacant lots in the City of Chicago at the census tract level?” A census tract is the smallest unit of analysis for which the United States government provides the income data relevant to such analysis; we decided this level of granularity would provide the most informative and relevant visual relationship. We also collect ZCTA area measures and perform a crosswalk to aggregate tract data up to the larger areas.

In order to do this, we first made a call to the US Census API using an individual key code. Using this key we downloaded Chicago tract-level data for the years 2010 through 2022 on: the number of households by income bracket and the number of housing units by market value bracket. That is, how many households in each tract fell into one of several income or home-value brackets.

We wrote this code in such a way that it prevents the user from having to download the required data every time they render the Quarto document. An if/else statement tells the computer to check if the needed data is present locally. If not, it downloads the data and names it appropriately. If so, it loads the local data for use in the next step.

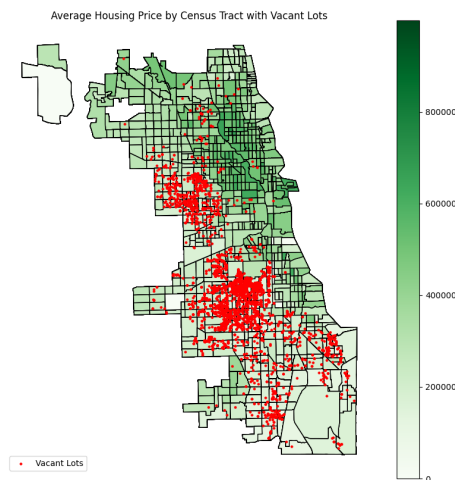
This data was then cleaned and organized to be easy to work with and render spatially after being combined with corresponding geo-spatial data. The raw data from the Census does not feature human-readable column names and these have to be pulled from yet another government website.

After this, we pulled data from the City of Chicago Data Portal that provided us with the location of the city's reported vacant lots, as well as the date of the vacancy violation. After some processing, we can then see a quick Altair mapping of that data, which lets us see the number of vacancy violations by year.



Now we can use longitude and latitude in the Chicago vacant lot data to build a geospatial dataframe. This then allows us to create a basic map of all of the reported vacant lots in the city. One difficulty with the vacancy data is that it only captures vacancies that have been registered with the city, indicating that there may be underreporting.

We combined our census data with geospatial data from the City of Chicago data portal, which allowed us to use matplotlib visualization grammar to show where vacant lots are most prevalent in the city in relation to the mean income and housing price in each Census Tract in Chicago. As a side note, we simply calculated the means from our census data.



And then we feed all of this code into our Shiny application! This app has two tabs, one that focuses on the location of aggregated lots and the area measure of wealth, and one that shows a map of yearly reports of vacant buildings.

The first tab allows you to choose between the two wealth measures (average area income) and two area measures (ZCTA and Census tract) using radio buttons.

When adjusting these two variables, we see that the vacant buildings are concentrated in areas with housing values between \$400,000 and \$200,000. Toggling to the income measure, the reported buildings cluster in tracts with average estimated household incomes between \$250,000 and \$750,000. We note that while vacant buildings are more concentrated in tracts with lower average estimated home values, and lower estimated household incomes, there are regions, specifically the far southeast side that have lower average estimated home values and lower estimated household incomes. This suggests that low home values and low household incomes are predicates not determinants of concentrated vacancies.

Another potential reason that the far southeast side does not have a high concentration of vacancies, is that it could be zoned differently from the west and southwest sides. For example, South Deering in particular has lower home values but very few vacancy violations. According to [Wikipedia](#), most of the area (80%) is zoned as industrial, natural wetlands, or parks. Future research could look at vacancy violations by zone to see if there are any distinct patterns.

The second tab of our Shiny app allows you to adjust the year displayed in this interactive map. While previous maps showed aggregate reported vacant buildings, this map adds in a temporal element, allowing us to see variation in reports for each additional year. Clicking on a point will give you the street address.

Future work will expand on this research by investigating a causal relationship between vacant buildings and area wealth measures using a linear regression.