

Graphic Detail

Geographic Information Systems (GIS) organize and clarify the patterns of human activities on the Earth's surface and their interaction with each other. GIS data, in the form of maps, can quickly and powerfully convey relationships to policymakers and the public. This department of Cityscape includes maps that convey important housing or community development policy issues or solutions. If you have made such a map and are willing to share it in a future issue of Cityscape, please contact john.c.huggins@hud.gov.

Visualizing Residential Vacancy by Length of Vacancy

Alexander Din
Dewberry

Data

The United States Postal Service (USPS) collects counts of occupied and vacant residential and business addresses across the United States. The counts of vacant addresses are broken down by length of vacancy. This information is useful for researchers, planners, analysts, and others concerned with vacancy issues in making informed decisions to address them. For example, communities affected by recent vacancy may require different approaches and solutions than communities affected by long-term vacancy. I demonstrate how to use a modified box plot with swarm plot, paired with a micromap, to visualize ratios of different lengths of residential vacancy compared with total residential vacancy. I developed visualizations at the census tract level for the Pittsburgh, Pennsylvania Core Based Statistical Area (CBSA) on a quarterly basis from the first quarter of 2012 until the first quarter of 2017.

The U.S. Department of Housing and Urban Development provided the USPS data in tabular format, and the data were joined to spatial census tract data using ArcGIS 10.4, a geographic information system. The Pittsburgh CBSA is composed of 711 census tracts. Each census tract has counts of residential vacancies for four categories: less than 1 year,¹ between 1 and 2 years, between 2 and 3 years, and greater than 3 years. I created ratios for each category by dividing the count for that category by the total number of vacant residences; thus, one census tract produces ratios for each category that describe the composition of each census tract. It is important to note that the ratios do not indicate whether residential vacancy counts are low or high in a census tract or whether a low or high percentage of residential addresses are vacant within the census tract.

¹ Counts of residential vacancies of 1 year or less were transformed from several fields in the original data set, which were vacancies of 3 months or less, 3 to 6 months, and 6 to 12 months.

Modified Box Plot With Swarm Plot

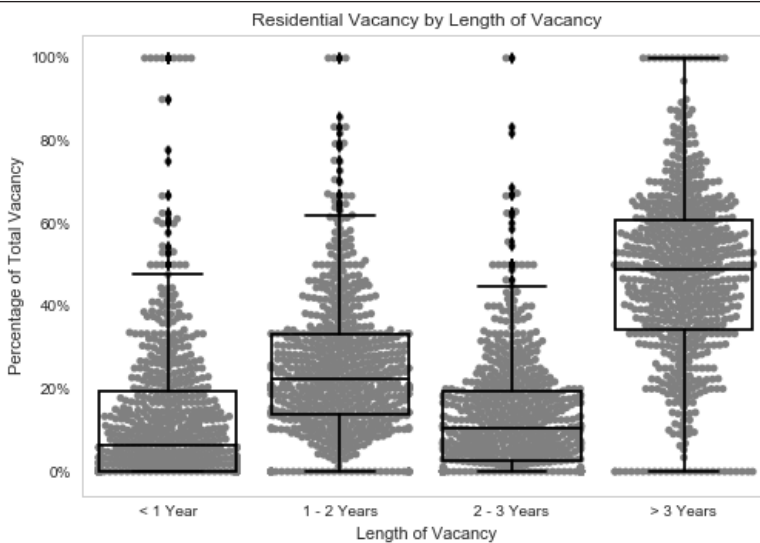
In order to understand each of the four categories for 711 census tracts, I developed a set of descriptive statistics to more succinctly represent the data set. The five-number summary is a set of descriptive statistics that includes the minimum, lower quartile, median, upper quartile, and maximum values of a data set. The standard method for visualizing the five-number summary is the box plot, also called a box and whiskers plot. At the center of the box plot is the median value; from there, the box extends to the lower and upper quartiles, and the whiskers extend to the minimum and maximum values of the category. However, the minimum and maximum values are not always the smallest and largest values but are instead capped at 1.5 times the interquartile range. Values outside of the minimum and maximum range are considered potential outliers and are represented with points. The outlier points for the standard box plot do not indicate whether multiple potential outliers exist at a specific value.

One limitation of the box plot is that it does not show the distribution of points within the plot. To overcome this limitation, the box plots have been modified to include a swarm plot. The swarm plot places a point for each census tract at the corresponding value along the box plot. If multiple census tracts have the same value, the swarm plot then plots the points outward horizontally, indicating a grouping of census tracts at a particular value, including at potential outlier values. This information is useful to visually inspect for potential clustering of values.

The graphic in exhibit 1 shows box plots for each category of residential vacancy for the first quarter of 2012 in the Pittsburgh CBSA. For residences that have been vacant 1 year or less, the median is slightly less than 10 percent. This indicates that in 50 percent of census tracts, residences that have

Exhibit 1

Box Plot Showing Distribution of Each Vacancy Type for the First Quarter of 2012



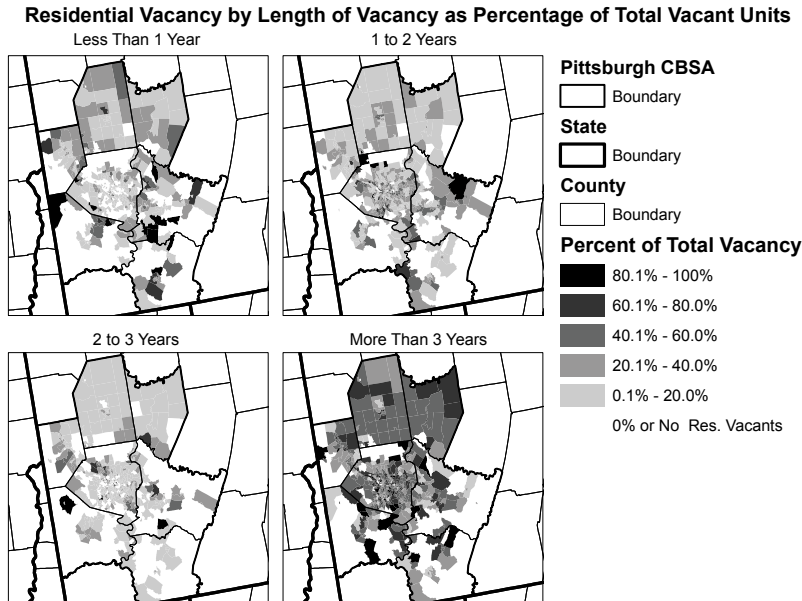
been vacant for 1 year or less compose less than 10 percent of all residential vacancies in those census tracts. For residences that have been vacant for 2 to 3 years, the maximum is approximately 45 percent. Census tracts where 45 percent or more of vacant residences have been vacant for 2 to 3 years may be potential outliers, including the grouping of census tracts at approximately the 50-percent mark. All categories except residences vacant for more than 3 years have potential outliers. The set of modified box plots present a clear method for visualizing the composition of residential vacancy across the Pittsburgh CBSA.

Micromap

The micromap is a set of maps that display related information on a single graphic (Mast, 2014). The micromap displays the four categories of residential vacancy using the same equal interval classification system in order to make the maps directly comparable to each other. Visualizing the data as a set of maps enables the reader to visually inspect the data in a spatial context. In exhibit 2, residences that have been vacant for less than 1 year appear to make up a greater proportion of vacant residences in the outer counties of the Pittsburgh CBSA. Residences that have been vacant for greater than 3 years appear to be concentrated in the central and northern regions of the Pittsburgh CBSA.

Exhibit 2

Micromap Showing Vacancy of Each Type for the First Quarter of 2012



Repeatable Process

For this analysis, I processed data and created outputs for 21 time frames. In exhibit 3, the median for residences vacant for less than 1 year increased to approximately 20 percent, and the median for residences vacant more than 3 years increased to nearly 60 percent. Coupled with the reduced range of the minimum and lower quartiles of the medium-term box plots, these changes indicate that recent vacancies and long-term vacancies compose a greater share of residential vacancies, whereas medium-term vacancies compose a smaller share. Comparing exhibit 4 with exhibit 2 shows that medium-term vacancy constitutes a smaller proportion of total residential vacancy across many census tracts, whereas recent vacancy and long-term vacancy appear to constitute a larger proportion.

The power of modern GIS and programming languages allows for the quick computing of multiple iterations of data processing and output techniques. Researchers, analysts, and others can quickly inspect visualizations across many time frames. While processing the data for these visualizations, one can transform data for further analysis, such as clustering, trending, and forecasting.

Exhibit 3

Box Plot Showing Distribution of Each Vacancy Type for the First Quarter of 2017

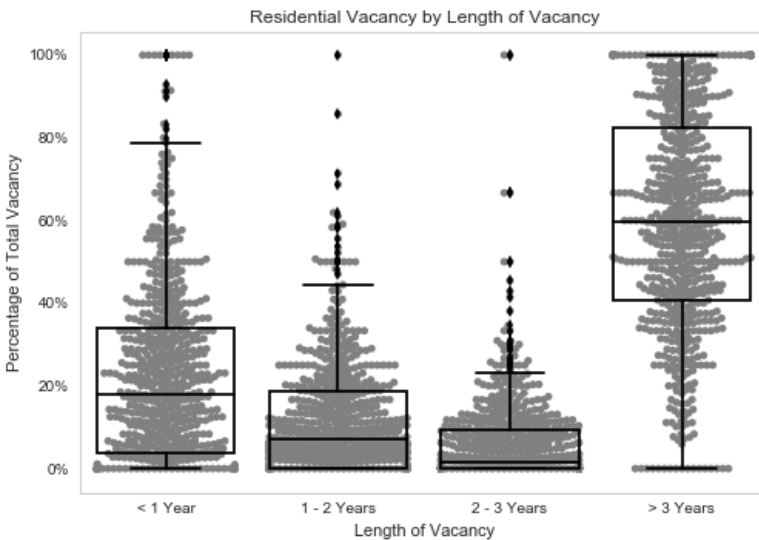
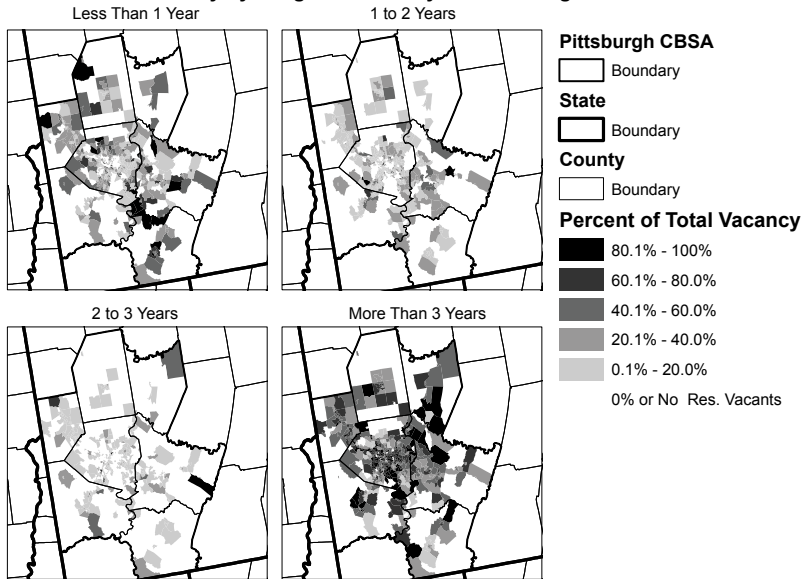


Exhibit 4

Micromap Showing Vacancy of Each Type for the First Quarter of 2017

Residential Vacancy by Length of Vacancy as Percentage of Total Vacant Units



Notes

ArcGIS 10.4 and the Spyder IDE were used for data processing. Python libraries used were Pandas 0.19.2, Matplotlib 2.0.0, and Seaborn 0.7.1. The Python scripts are available at <https://github.com/alexdingis/residential-vacancy-by-length>.

Author

Alexander Din is a Geographic Information Systems analyst at Dewberry.

References

Mast, Brent. 2014. "Comparative Micromaps and Changing State Homeownership Rates," *Cityscape* 16 (2): 163–167.