

Visualizing and Comparing Housing Values in Major Metropolitan Cities

Sophia Zheng

12/15/2019

Motivation

The primary question motivating my project is how housing prices can be compared among three major metropolitan areas in the United States, namely the New York, San Francisco, and Seattle Metropolitan Areas. My interest was particularly drawn to these three regions because of my hypothesis that housing price variations are highly correlated with income changes. I wanted to look at two historically expensive regions, New York and San Francisco, to see what might be predicted about prices in Seattle in the next couple of years. It is particularly interesting that the San Francisco area has faced increasingly high housing prices over the past thirty years, which I attribute to the influence of Silicon Valley. This led to my hypothesis that Seattle is posed to be the next major metropolitan area to experience a big jump in housing prices, similar to that of San Francisco, although perhaps not to quite the same extent. This hypothesis is also rooted in the tech industry, as big technology companies take hold of more and more of the city and increase the average incomes of the city's residents along the way.

Methodology

The dataset for this project was collected from the online real estate database company, Zillow at this link. I acknowledge that this dataset might not be the most representative of the true state of housing prices in the US, since this is only one company's compiled data - however, I have chosen to use this data, despite possible skews in the data, as Zillow is one of the largest online real estate companies in the US, and has collected a robust set of data points over the past thirty years.

The measures that I have used include the Zillow House Value Index (ZHVI), sale prices, median sale list prices by square foot, Zillow Rent Index (ZRI), median rental prices, median rental list prices by square foot, and median income. I have focused more on the ZHVI and the median sale list price by square foot, since the rental measures are less relevant as short-term housing options.

With the data, I have explored the data in three ways - firstly, by comparing the three regions over time utilizing a number of different aggregate measures; secondly, by looking at the relationship between median income and average housing value in the three regions; and thirdly, by looking at the maps of the areas, specifically comparing housing value within each metropolitan area between zip codes.

Discussion

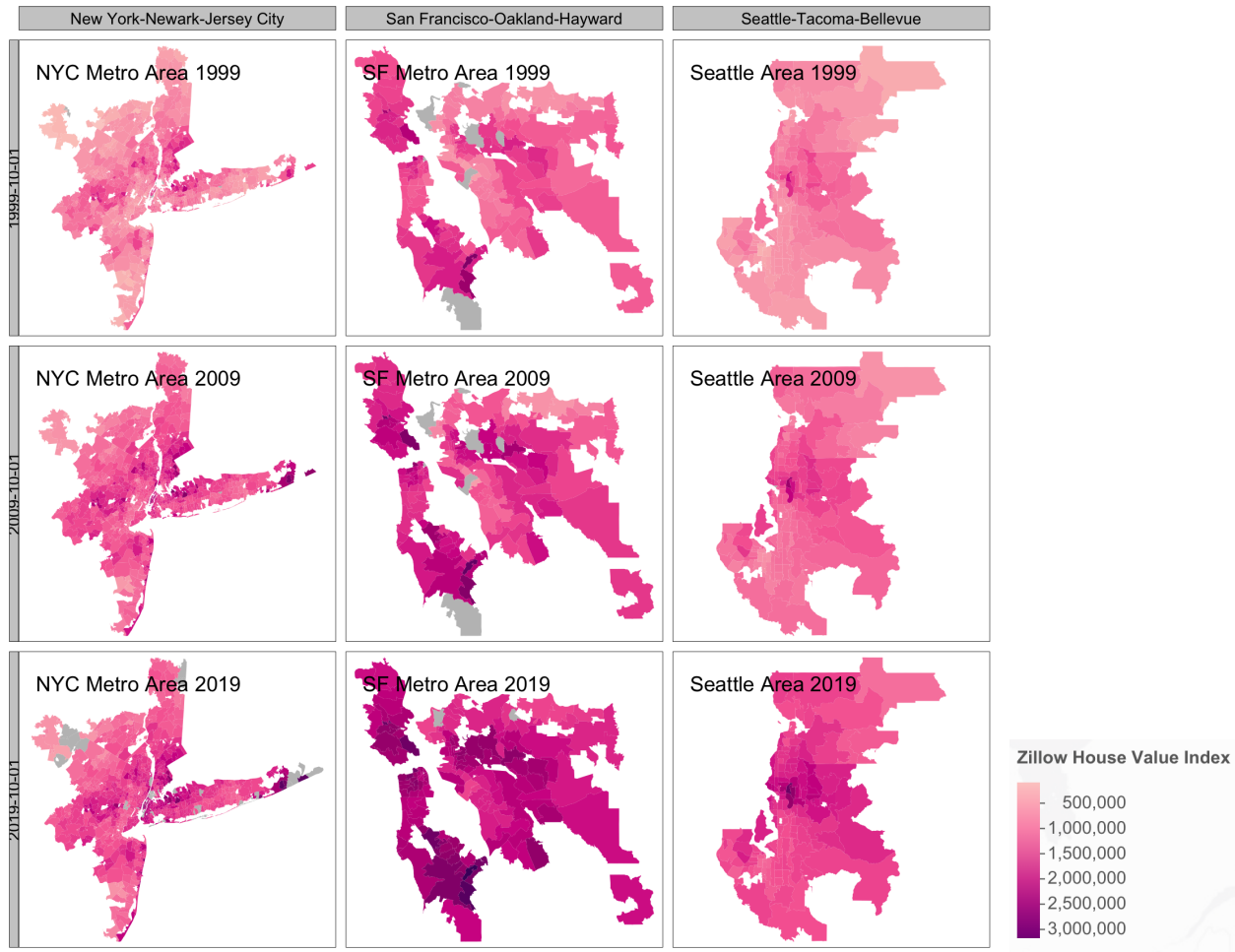
The first way that I explored the data was through a series of graphs comparing various factors among the three regions. Looking first at the ZHVI graph, we can see that the line for Seattle has been increasing faster than New York since around 2012, and that the housing values in Seattle in 2019 have just begun to surpass those of New York. When looking at sale prices, this distinction is especially noticable, with the average home in the Seattle metropolitan area selling for more than the average New York home since the end of 2015. Of course, this is not normalized, so when we look at the sale list prices per square foot, we see that some of the higher prices in Seattle are simply due to larger homes. However, Seattle's list prices per square foot have been increasing at a slightly faster rate than New York's since 2012, so if the trends continue, the lines will likely intersect in the near future. I also looked at rent prices - however, the correlations aren't quite as clear with these measures, given that the nature of rentals is for short term housing.



After doing this first exploration of the data, I wanted to look more closely at the correlation between housing prices and income, so I plotted income over ZHVI and animated this graph over time. I found that this graphic was particularly useful in visualizing the relationship between the two measures over time. My expectation was that housing value and income would be positively correlated, since income is of course correlated with willingness-to-pay, and therefore housing prices would tend to increase with income. Therefore, I expected to see increases in income (in my graphic, shifts to the right on the x-axis), to be followed after some time with increases in ZHVI (shifts upward on the y-axis). The data for Seattle somewhat confirms this hypothesis - in 2012, Seattle's median income is the same as New York's, although New York has much higher average ZHVI. However, in recent years, Seattle's average ZHVI has increased rapidly to match New York's average.



Finally, I decided to look more closely at the distribution of housing values within each metropolitan area. For this, I made maps of the three regions divided into separate zip codes, and colored them based on the ZHVI. I did this for data points in 1999, 2009, and 2019 to see change over time. Interestingly, it seems that the San Francisco area has over the past thirty years had high ZHVI not only within the city itself, but also extending into the surrounding area, whereas New York's high value homes have mainly been concentrated within New York City proper, while not really spilling into the New Jersey and Pennsylvania portions of the NY metro area. Seattle, by comparison, while not as extreme as the Bay Area, has high house values across the entire Seattle metropolitan area, not just within the city.



Areas of Further Exploration

This project could be extended in a number of ways for more concrete conclusions. For example, one large determiner in housing prices is population, so one path that I considered exploring is looking at the population changes within each metropolitan area. Another interesting exploration would combine thd income analysis with the zip code breakdown, and would look to see the correlation between income and home value within each zip code. However, I was unable to find a dataset with this information, as it is not readily available from the census data.