An Exploratory Analysis of the Relationship between Housing Prices and Evictions in the U.S.

W200 Fall 2018 Project 2 | Adam Sohn & Paul Petit

Introduction

Aside from being a prudent way to build wealth over time, home ownership has long been seen as a centerpiece of the American dream. According to the U.S. Bureau of the Census, roughly 65%[1] of households are owner-occupied. However, that figure may be trending downward in the near future[2] as nationwide home affordability concerns regularly garner headlines of popular media sources.

The presumed upshot of a decrease in home ownership is an increase portion of renters. However, in a similar parallel trend, news outlets have reported an increase in rate of eviction[3], a devastating experience for tenant and significant financial liability for landlord.

In this project, we analyze the relationship between home affordability, eviction, wage, time, and other variables. In what follows, we set a guiding list of questions and test our economic intuitions (our hypotheses) in these markets to glean insights that make sense of a rich nationwide housing price and eviction dataset.

Questions

1. How does regional home affordability relate to regional rental affordability?
2. How does eviction rate change with cost of rent?

---

[1] "Homeownership Rate for the United States." FRED, Federal Reserve Bank of St. Louis, 30 Oct. 2018, fred.stlouisfed.org/series/RHORUSQ156N#0.

[2] Kishan, Hari. "U.S. House Prices to Rise at Twice the Speed of Inflation and Pay:..." Reuters, Thomson Reuters, 6 June 2018, www.reuters.com/article/us-usa-property-poll/u-s-house-prices-to-rise-at-twice-the-speed-of-inflation-and-pay-reuters-poll-idUSKCN1J20G3.

[3] McMullen, Troy. "Why More Americans Are Facing Eviction." Forbes, Forbes Magazine, 12 Dec. 2016, www.forbes.com/sites/troymcmullen/2016/12/12/why-more-americans-are-facing-eviction/#33bf1d911617.

Data Sources

Eviction Data

Our primary dataset contains eviction data by year and geography. This dataset also conveniently includes demographic and income fields like population, median household income, and median gross rent by year and geography, which will allow us to test more sophisticated hypotheses. Further details in the table below.

EVICTION DATA	
Source	https://data-downloads.evictionlab.org/
Description	Data collected by the Eviction Lab, a research organization dedicated to studying the prevalence, causes, and consequences of eviction, including census data. The data is comprised of formal eviction records from 48 states and the District of Columbia from 2000 through 2016.
Grain	Year || Metropolitan Area [City] || parent location [State]
Size	67.6 MB unfiltered (502,872 rows, 27 columns) at finest grain

Field of Interest	Description	Sample Value	Notes
population	Total population	100,000	Used for filtering
eviction rate	Ratio of the number of renter-occupied households in an area that received an eviction judgement in which renters were ordered to leave	0.32	Field for analysis for second hypothesis
median household income	Median household income	63,287.00	Field of analysis for both hypotheses
median gross rent	Median gross rent monthly	1327.65	Field for analysis for both hypotheses

Housing Price Data

We adjoined to our primary dataset an additional dataset with that contains residential real estate data and is described below.

RESIDENTIAL REAL ESTATE DATA	
Source	https://www.realtor.com/research/data/.
Description	Data contains residential (single family home and condo/townhome) real estate inventory data from May of 2012 through October 2018 at various geographic grains for the U.S. According to the website, "Data in this realtor.com library is based on the most comprehensive and accurate database of MLS-listed for-sale homes in the industry. We aggregate and analyze data from hundreds of sources and produce hundreds of metrics for multiple markets, and curate figures and trends where possible for reliability and comparability."
Grain	Year || Metropolitan Area [City] || County || State
Size	10 MB (39,001 rows, 35 columns) at finest grain

*Items in brackets (ie. [YoY]) are calculated fields that represent t_1 - t_0

Field of Interest	Description	Sample Value	Notes
Median listing price	Median listing price	452,600.00	Field for analysis for first hypothesis

Data Cleaning and Assumptions

Our data cleaning approaches and the assumptions we made were analysis-dependent. We describe below our data cleaning approach and assumptions by question/hypothesis number:

1. Key themes in the data cleaning and assumptions setting phase were to ensure that data is present, rational, and that datasets could merge cleanly.

a. Some city names occur in more than 1 state. This created a non-value-added complexity for analysis, so data was narrowed down to cities with population > 50,000 to eliminate duplicates across focus cities.
b. To take advantage of the richness of both datasets, a join was required.
    i. The primary concern was ensuring a valid join index. This was a challenge as eviction dataset described areas by city ('name') and realtor dataset described areas by 'CountyName'. Thus, only cities wholly owning 1 county were applicable. Making this match was a manual process due to a lack of compatible dataset matched on nomenclature. This manual process limited our location count to 20 cities, as below.

	name	CountyName
0	Detroit	Wayne, MI
1	Atlanta	Fulton, GA
2	San Francisco	San Francisco, CA
3	Newark	Essex, NJ
4	Denver	Denver, CO
5	Las Vegas	Clark, NV
6	Tulsa	Tulsa, OK
7	Seattle	King, WA
8	El Paso	El Paso, TX
9	St. Louis	St. Louis, MO
10	Chicago	Cook, IL
11	Miami	Miami-Dade, FL
12	Cleveland	Cuyahoga, OH
13	Los Angeles	Los Angeles, CA
14	Dallas	Dallas, TX
15	Minneapolis	Hennepin, MN
16	Boston	Suffolk, MA
17	San Diego	San Diego, CA
18	Philadelphia	Philadelphia, PA
19	Orlando	Orange, FL

    ii. This analysis focussed on 2016 as the data year, as it is the most recent data that existed in the joined dataset. As there was difficulty finding a data set for matching counties to cities, the joined dataset needed to be joined on a decoder ring of a limited number of cities that wholly contain a single county.

2. To examine the relationship between changes in median gross rent and changes in eviction rate, we relied on the eviction dataset.
    a. Initial exploratory analysis of the data revealed that we needed to include parent location (state) in addition to metropolitan area because there are multiple metropolitan areas across the U.S. with the same name (eg. "Portland, AK", "Portland, CO", "Portland, OR"; "Oakland, CA", "Oakland, FL", "Oakland, IL").
    b. Missing data for small metropolitan areas and/or certain years forced us to consider filtering our dataset. Additionally, given the analysis was looking at trends across time, small metropolitan areas, even with all data

accounted for, witnessed high volatility in their YoY % changes in median household income and median gross rent figures. As such, we made the decision to filter our dataset to include metropolitan areas with a population of 100,000 or more. Filtering in this way left us with 564 data points (282 per year change).

c. Though median household income and median gross rent variables should be continuous across time (they should change year-over-year), our dataset was limited by the fact that these numbers were discrete across roughly 5-year (quinquennial) increments of time. There was one exceptional year (2010), however, when the data were updated such that these variables were continuous between 2009-2011, which allowed us to make the observations we wanted to over a sufficiently large time window. The limitation with this approach is that our finding for the relationship between changes in household income, gross rent, and eviction rate are likely not externally valid given peculiarities about this particular time window (ie. 2009-2011 was in the wake of the recession, and so, our findings from this time period may not generalizable across time).

Analysis

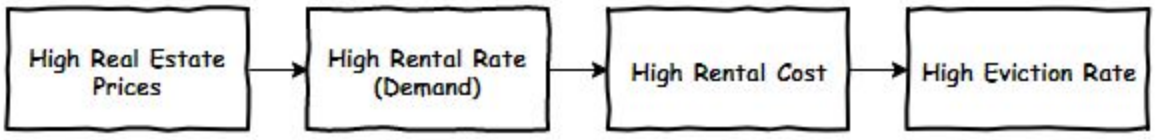
Hypothesis 1

Our first question asks, “How does regional home affordability relate to regional rental affordability?” Hypothesis 1 answers this question with:

Regions where home ownership is relatively expensive will be relatively less affordable for renters. This will manifest in a higher rates of rental and rental eviction rates.

The hypothesis follows the argument made by advocates of affordable housing; high real estate prices drive the rental market, which ultimately will price people out of their rentals.

Figure 1.1 Affordable Housing Continuum



We begin the analysis by examining the variance in median listing price and median household income across 20 major US cities as seen below in Figure 1.2.

Figure 1.2 Real Estate Prices

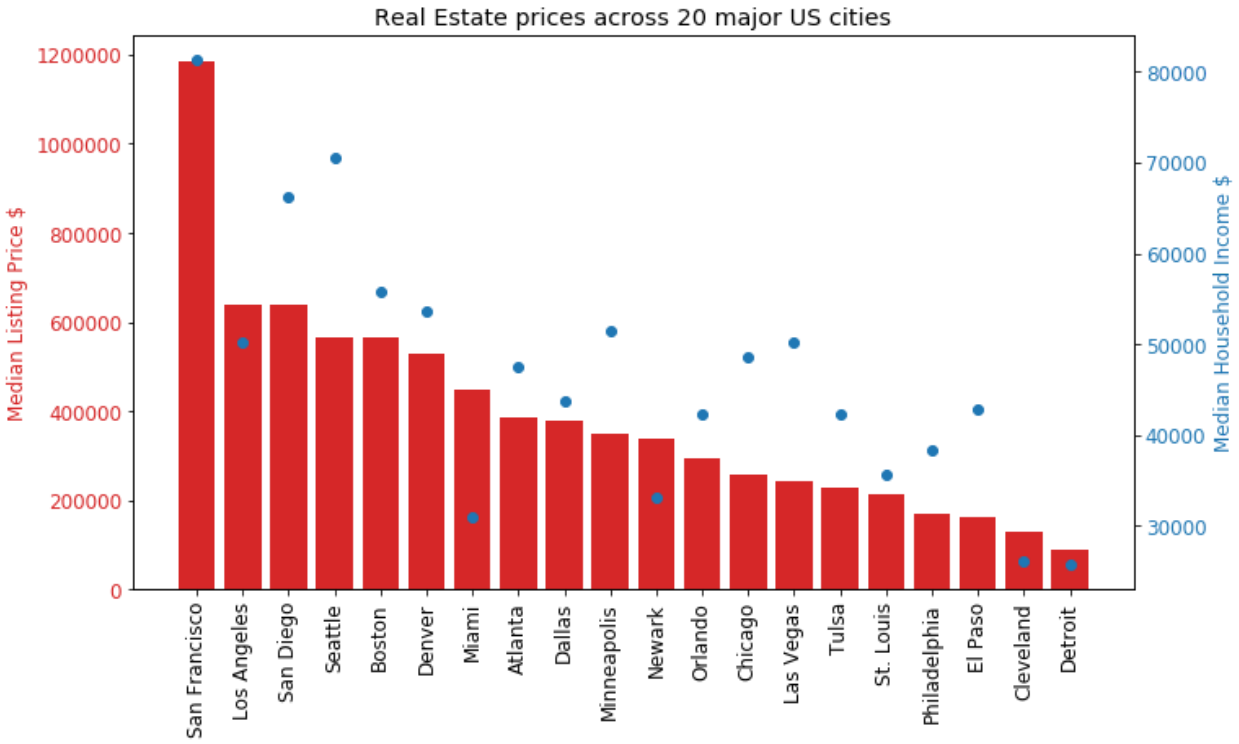
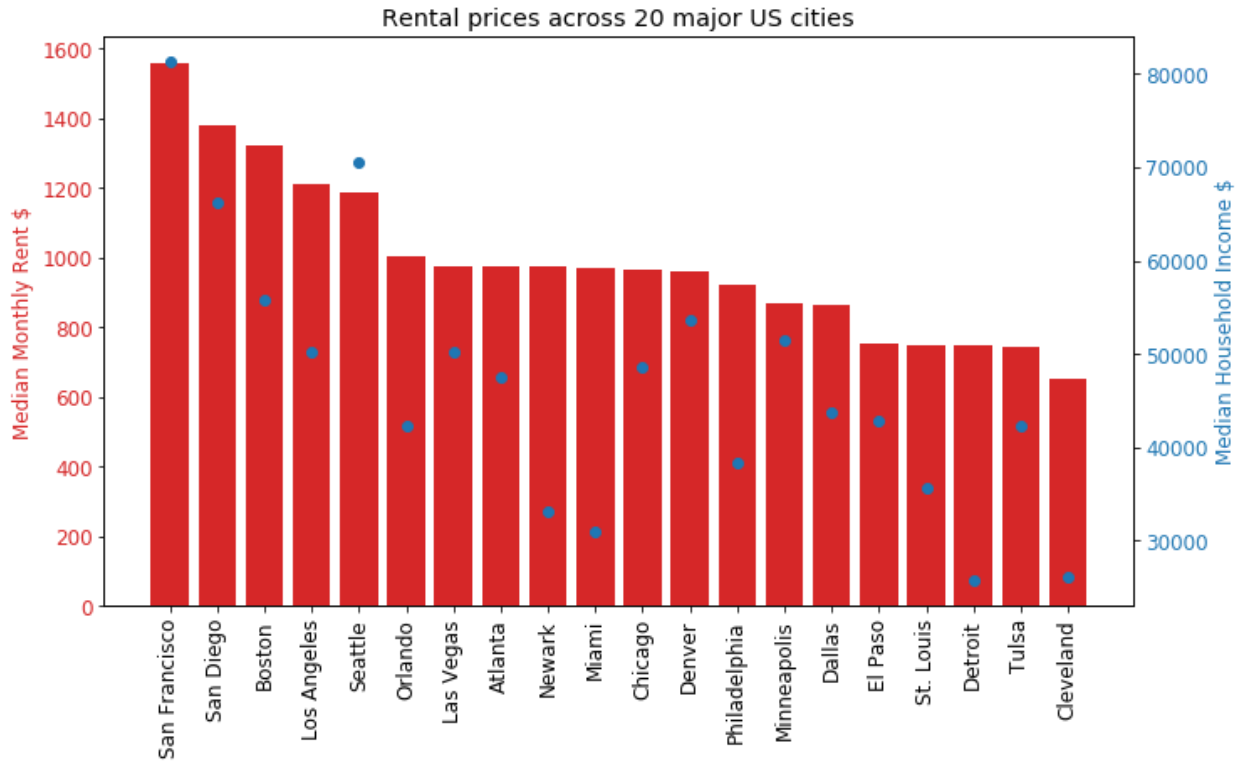


Figure 1.2 shows a dramatic disparity across the selected cities for median listing price and median household income. However, with the exception of San Francisco and Miami, the two variables appear to be following a similar trend.

Rental prices have a similar relationship to median household income as seen below in Figure 1.3.

Figure 1.3 Rental Prices



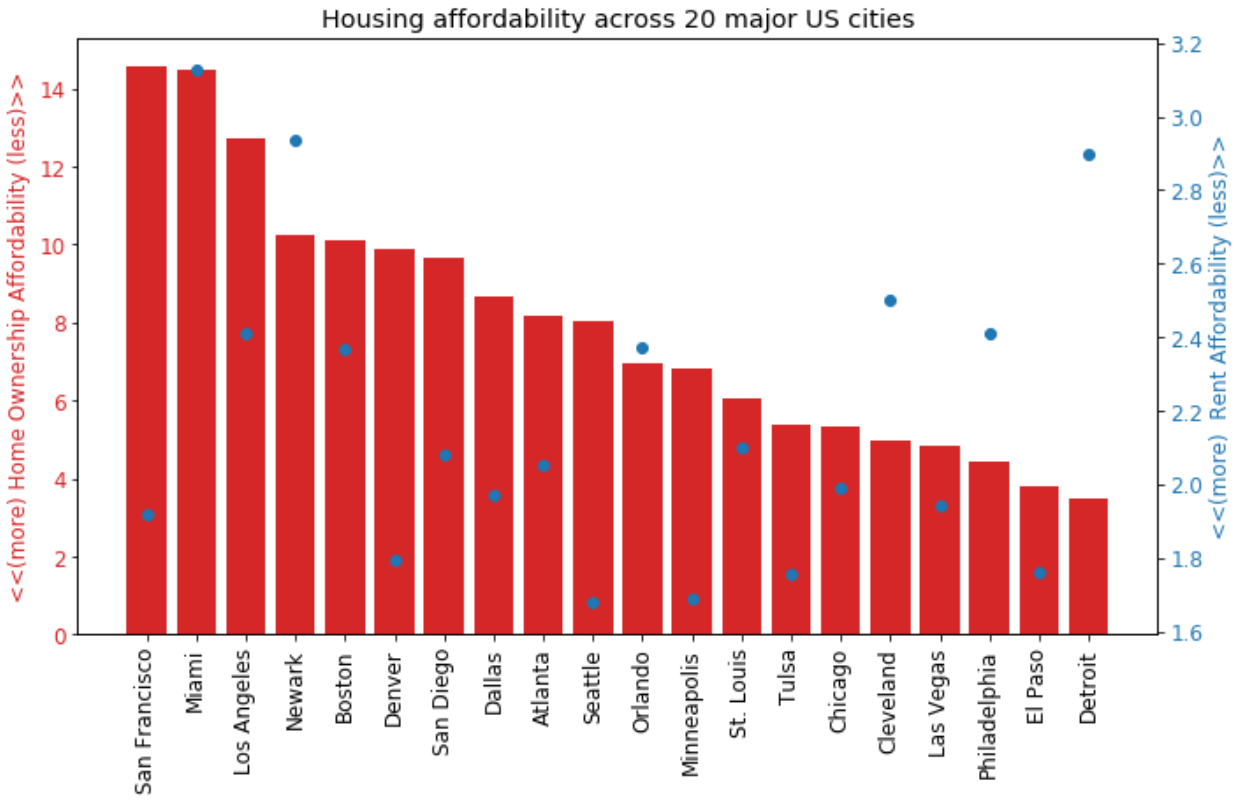
Rental prices appear to be a flatter trend relative to income. Note that for Figure 1.3 and future graphs of the same type, cities on the x-axis are resorted to maintain a declining trend.

To further the analysis by examining rates of rental, it is important to condense the above information into affordability metrics45 as described below.

> Home Ownership Affordability = Median Listing Price / Median Household Income
> Rent Affordability = 100 * Median Monthly Rent / Median Household Income

Figure 1.4 below shows the affordability metrics.

Figure 1.4 Affordability



When looking at Figure 1.4, there does not appear to be a strong relationship between affordability of Home Ownership Affordability and Rent Affordability. This serves to refute the first part of Hypothesis 1:

Regions where home ownership is relatively expensive will be relatively less affordable for renters...

There are some cities that maintain the hypothesized relationship (ie. Miami, St. Louis, El Paso), however more that are enigmatic (ie. San Francisco, Seattle, Detroit).

San Francisco's rent affordability data is out of line with the volume of anecdotal noise regarding rents in the $3k / month range. However, the explanation is that 60% of San Francisco renters enjoy rent control to keep rents lower than market price.[4] Also, incomes are high in the Bay Area, which modulates the affordability metric.

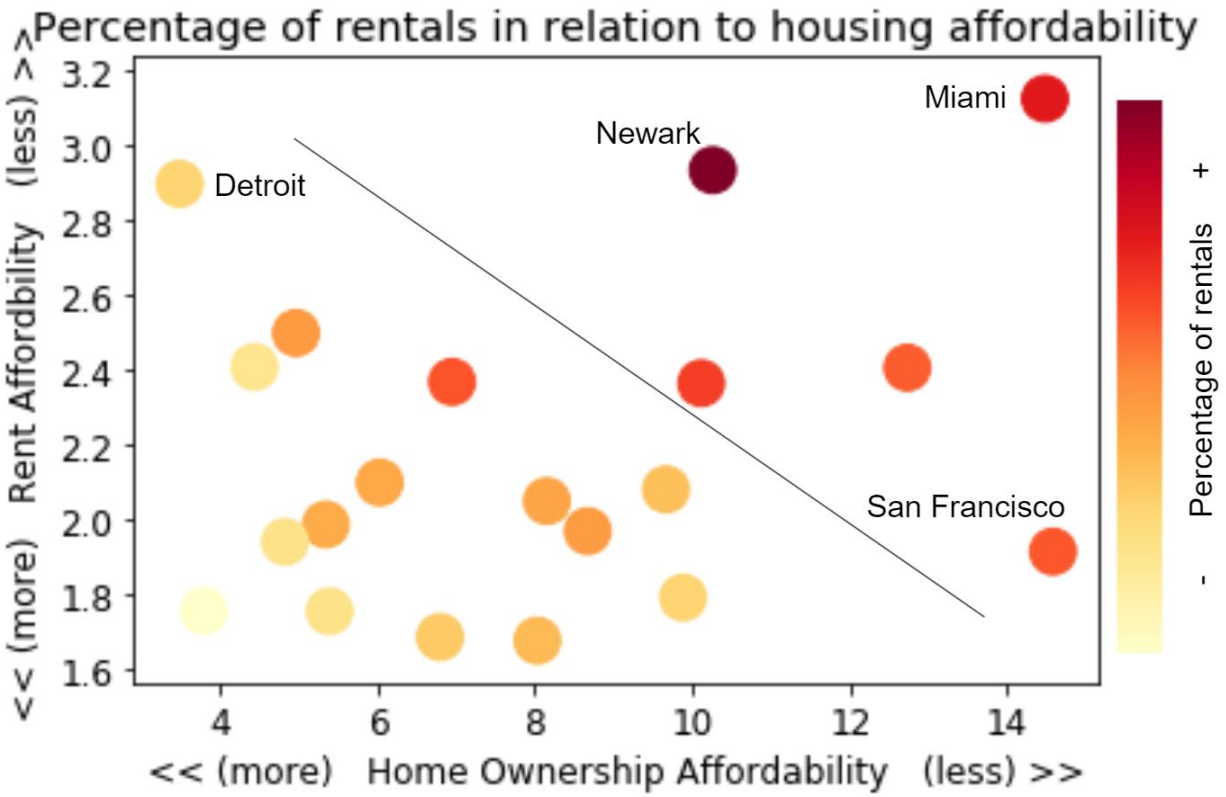
[4] Brinklow, Adam. "More than 60 Percent of SF Renters Have Rent Control, Says City." Curbed SF, Curbed SF, 12 July 2018, sf.curbed.com/2018/7/12/17565192/housing-needs-trends-report-rent-control-san-francisco#.

Seattle's rent affordability is a product of the high wages in the city, based on several large tech companies with operations in the area.

Detroit is on the other side of the affordability spectrum. Real estate is far more affordable than rent. It would be expected that rent would also be affordable. However, Detroit Metro Times reports a large cost burden placed on landlords by recent code updates. These code updates cost between $4k - $9k per unit[5].

Moving on to the assertion that home ownership affordability issues will lead towards rental demand , we examine Figure 1.5

Figure 1.5 Rate of Rental



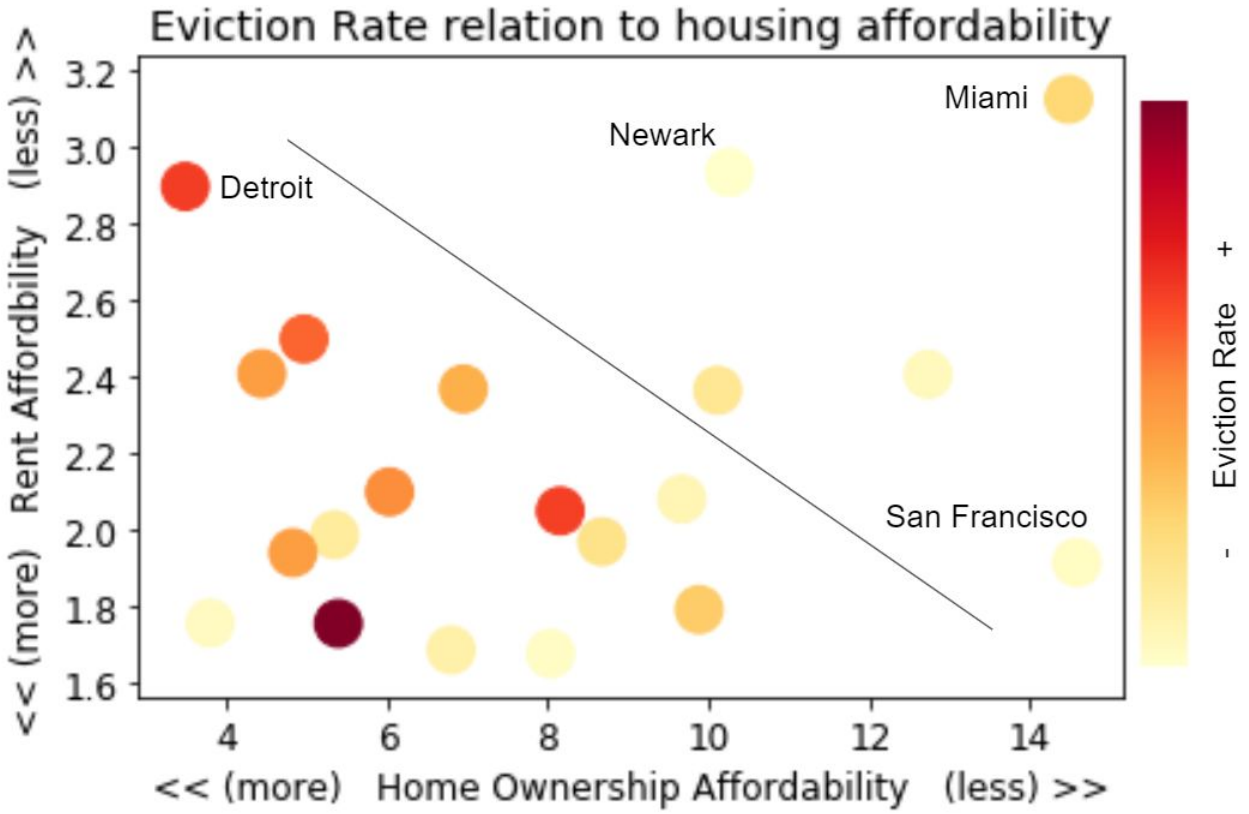
[5] Ikonomova, Violet. "Higher Rents, 'Massive Displacement': The Unknown Cost of Detroit's Landlord Crackdown." Detroit Metro Times, Detroit Metro Times, 14 Dec. 2018, www.metrotimes.com/detroit/higher-rents-and-massive-displacement-the-hidden-cost-of-detroits-landlord-crackdown/Content?oid=15129860.

Figure 1.5 shows that areas that are less affordable for owning a home do have higher higher rental rates. This supports the underlined portion of Hypothesis 1:

Regions where home ownership is relatively expensive will be relatively less affordable for renters. This will manifest in a higher rates of rental...

The final portion of the analysis for Hypothesis 1 test the axiom of less affordable rents leading to higher eviction rates.

Figure 1.6 Eviction Rate



The finding here is surprising. It is in the more affordable regions where eviction rates are higher. The data does not contain an explanation for this, nor does other published works. This may truly be the fascinating finding of the study. To be clear, this further refutes the underlined portion of Hypothesis 1:

Regions where home ownership is relatively expensive will be relatively less affordable for renters. This will manifest in a higher rates of rental and rental eviction rates.

Hypothesis 2

Our second question asks, “How does eviction rate change with changes in cost of rent?” Hypothesis 2 offers the intuitive answer:

When cost of rent changes, eviction rate changes proportionately in the same direction.

Common economic reasoning would suggest that cost of rent is a primary factor in a tenant’s ability to afford to stay in their rental unit. If that cost of rent increases, then tenants are less likely to be able to afford their rent, and hence, more likely to be evicted. So to translate this into terms that we can observe in our dataset, we’d expect there to be a positive correlation between eviction rate YoY and median gross rent YoY across locations.

Note that for this hypothesis, the key relationship of interest isn’t simply between eviction rate and median gross rent, but between *year-over-year changes* in these variables. The hypothesis is not suggesting that locations with with low rent will likewise contain fewer evictions or that locations with high rent will contain many evictions — that belies intuition. Rather, it’s the velocity of these figures that we’re interested in.

Finally, two prefatory comments about the figures that follow.

1. The aforementioned (Data Cleaning and Assumptions, section 2c) limitation in our eviction dataset that resulted in median gross rent and median household income updating roughly quinquennially is relevant for this analysis. Because of this, we don’t have robust YoY data for our entire dataset. However, luckily, the quinquennial update pattern doesn’t apply from 2009-2011. Over this time frame, the data updated annually — meaning we have YoY median gross rent and household income data in 2010 and 2011 — which allows us to observe the relationship between eviction rate change and cost of rent change, as well as cut by household income (as we’ll do later in this analysis).
2. To minimize outliers that convolute findings and adopt best practice when analyzing economic data by location, the following figures are limited to metropolitan areas in the U.S. with a population of 100,000 people or more.

Figure 2.1

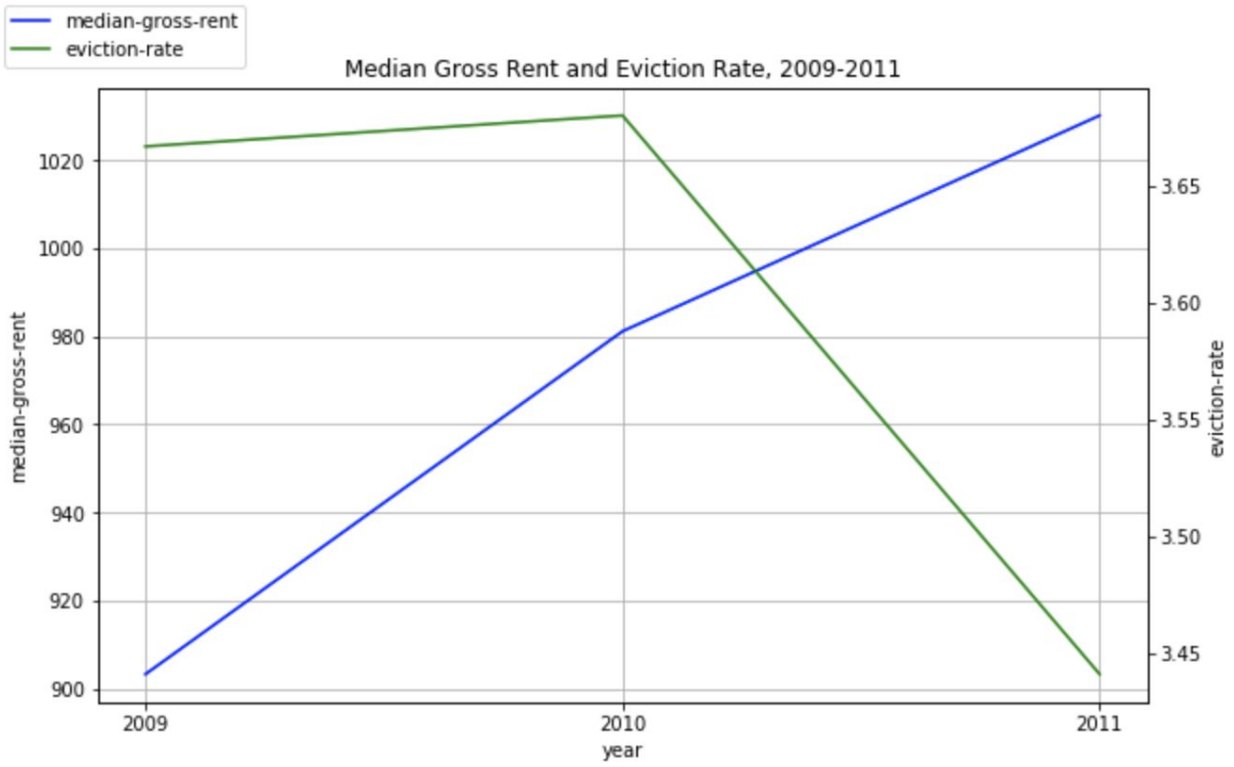
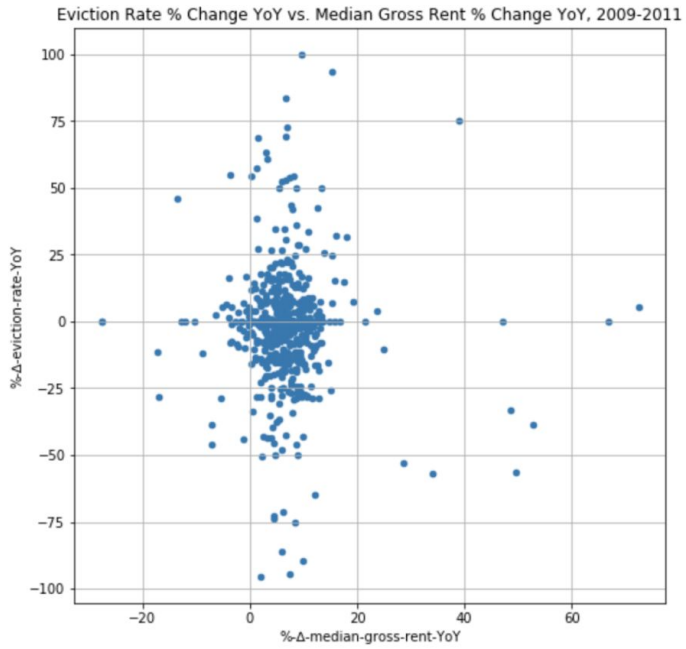


Figure 2.1 provides a high-level view into the YoY changes in median gross rent and eviction rate between 2009 and 2011. While median gross rent seems to be steadily increasing over this period, eviction rates increase slightly between 2009 and 2010, but them drop noticeably between 2010 and 2011. Given median gross rent is increasing over this time frame, this drop is a challenge to the credibility of our hypothesis.

To take a closer look at the relationship between YoY median gross rent changes and YoY eviction rate changes, we'll use a scatter plot to map individual cases of YoY rent changes and YoY eviction rate changes to look for the positive correlation our hypothesis posits in figure 2.2.

Figure 2.2



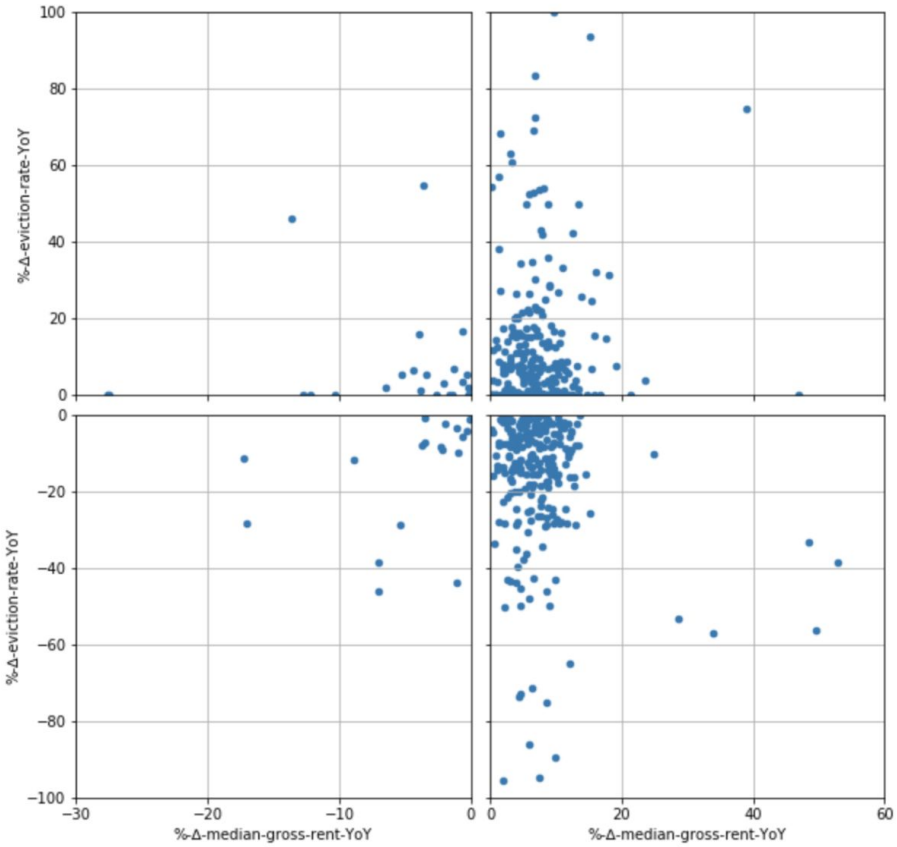
	%-Δ-median-gross-rent-YoY	%-Δ-eviction-rate-YoY
count	563.000000	563.000000
mean	6.454951	-1.945748
std	7.632659	21.974647
min	-27.622842	-95.522388
25%	3.327228	-10.156951
50%	6.033789	0.000000
75%	8.868411	5.377174
max	72.340426	100.000000

Figure 2.2 plots a point for each metropolitan area-year combination between 2009 and 2011. At a glance, the cluster of points around ~(6%, 0%) stands out. The variance in eviction rate also stands out, but the balance across the positive and negative side of the change in eviction rate is noteworthy given our expectation, per hypothesis 2, that we'd see a positive correlation between these YoY median gross rent and YoY eviction rate.

Given the positive correlation between eviction rate change and cost of rent change didn't emerge, **we feel it's safe to reject hypothesis 2**. When cost of rent changes, eviction rate **does not** necessarily change proportionately in the same direction.

To take a more careful look at this plot, especially given the surprise distribution, and to move our thinking forward, we've broken the plot into quadrants and described the data in each quadrant in figure 2.3 below.

Figure 2.3



Data Descriptions by Quadrant

	%-Δ-median-gross-rent-YoY	%-Δ-eviction-rate-YoY
count	23.000000	23.000000
mean	-6.328532	7.645103
std	7.861317	14.344970
min	-27.622842	0.000000
25%	-8.389279	0.000000
50%	-3.595506	2.185792
75%	-1.319929	6.101127
max	-0.137363	54.794521

	%-Δ-median-gross-rent-YoY	%-Δ-eviction-rate-YoY
count	276.000000	276.000000
mean	7.398456	11.152033
std	7.345263	17.288813
min	0.000000	0.000000
25%	4.010642	0.000000
50%	6.324232	5.158382
75%	9.127070	14.384010
max	72.340426	100.000000

	%-Δ-median-gross-rent-YoY	%-Δ-eviction-rate-YoY
count	18.000000	18.000000
mean	-4.660519	-15.044291
std	5.207397	15.045047
min	-17.273431	-46.292948
25%	-6.602533	-24.305556
50%	-2.883722	-8.768190
75%	-1.097211	-4.781171
max	-0.104822	-0.877193

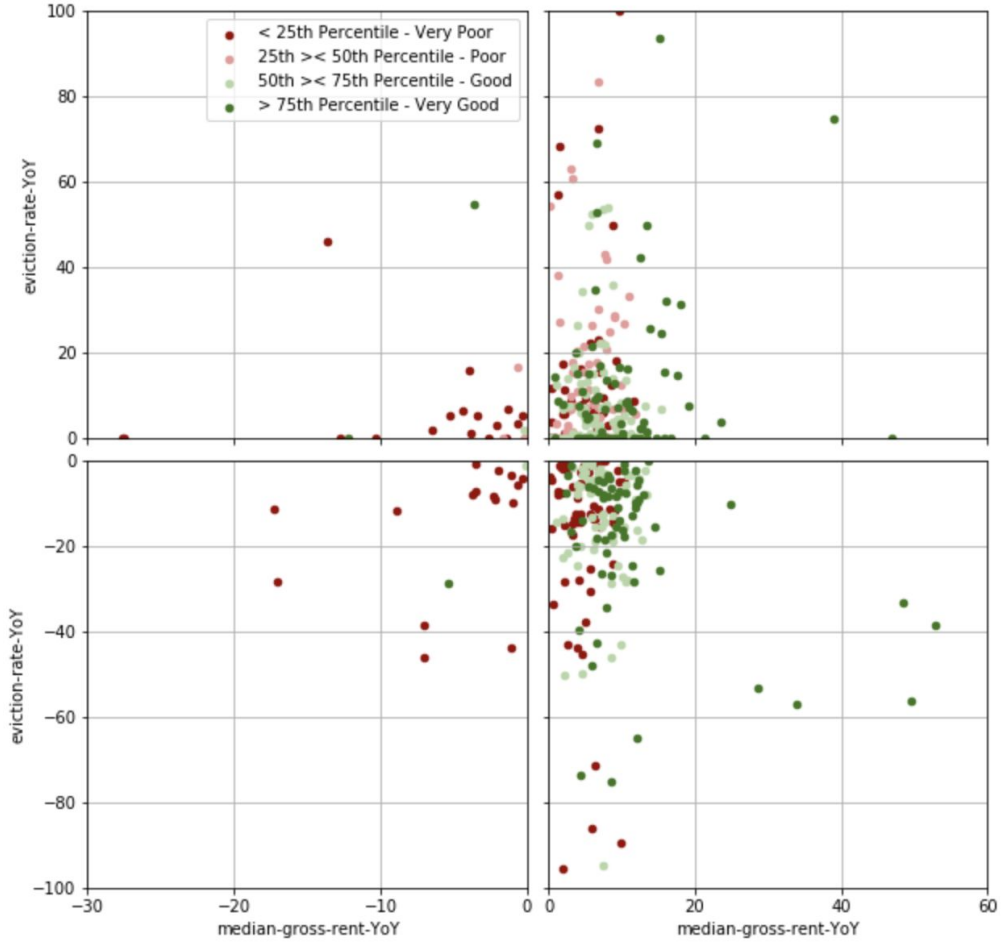
	%-Δ-median-gross-rent-YoY	%-Δ-eviction-rate-YoY
count	246.000000	246.000000
mean	7.404915	-16.579095
std	6.270373	17.563708
min	0.210970	-95.522388
25%	4.242948	-20.872394
50%	6.383853	-11.583184
75%	8.889968	-5.061948
max	52.803738	-0.161812

These quadrants in figure 2.3 allow us to see that in 93% of cases (522/563), YoY median gross rent increased in this 2009-2011 window. And when it increased, there were roughly the same number of cases in which eviction rate increased (53%) as there were cases in which it decreased (47%). This even distribution is the key finding that busts our hypothesis for this set of data. Furthermore, the distributions across the top right and bottom right quadrants show that when YoY median rent increased, YoY eviction rate decreased by more (mean of -17%) for whom eviction rate decreased than it increased by (mean of 11%) for whom eviction rate increased.

These two findings lend even more credence to the theory that **the correlation between median rent velocity and eviction rate velocity is not positive**.

To add a final layer of color to this investigation, we'll adjoin YoY median household income data as a categorical variable to our scatter plot. YoY median household income is categorized according to its quartile in the data series. So, within our set of data, a median household income change within the 25th percentile will show up as dark red as it constitutes a very poor YoY median household income change on a comparative basis. Changes between the 25th and 50th percentile (poor) will show as light red, changes between 50th and 75th percentile (good) will show as light green, and changes beyond the 75th percentile (very good) will show as dark green. (For specific bounds used as increments for these ranges, see the note beneath figure 2.4.)

Figure 2.4



Data Descriptions by Quadrant

```
median-household-income-YoY-categories
< 25th Percentile - Very Poor     17
25th >< 50th Percentile - Poor     3
50th >< 75th Percentile - Good     1
> 75th Percentile - Very Good      2
Name: name, dtype: int64
```

```
median-household-income-YoY-categories
< 25th Percentile - Very Poor     48
25th >< 50th Percentile - Poor    71
50th >< 75th Percentile - Good    80
> 75th Percentile - Very Good     76
Name: name, dtype: int64
```

```
median-household-income-YoY-categories
< 25th Percentile - Very Poor     16
25th >< 50th Percentile - Poor     0
50th >< 75th Percentile - Good     1
> 75th Percentile - Very Good      1
Name: name, dtype: int64
```

```
median-household-income-YoY-categories
< 25th Percentile - Very Poor     60
25th >< 50th Percentile - Poor    66
50th >< 75th Percentile - Good    59
> 75th Percentile - Very Good     61
Name: name, dtype: int64
```

**Note, the specific increment bounds used to categorize very poor, poor, good, and very good YoY median household income changes were Δincome < -0.77% (25th percentile), -0.77% <= Δincome < +1.73% (50th percentile), +1.73% <= Δincome < 4.41% (75th percentile), and Δincome >= 4.41%.*

The first noticeable feature of this figure is the apparent positive correlation between YoY median gross rent and YoY median household income: when YoY income increases, so does rent, and vice versa. That correlation is intuitive enough, and as such, isn't particularly interesting.

The more interesting and unintuitive relationship that's observable upon closer look at the counts of each income change category in the figure tables is the positive correlation between median household income and eviction rate. With some quick arithmetic, the following table shows the percentage of cases in which eviction rate increased by household income change category.

median-household-income-YoY-categories	% Cases in which Eviction Rate Increased
< 25th percentile -- very poor	46%
25th >< 50th percentile -- poor	53%
50th >< 75th percentile -- good	57%
> 75th percentile -- very good	56%

The upshot of the table above is that **there were actually more cases in which a metropolitan area's eviction rate *increased* YoY for areas that had a good or very good median household income change YoY as opposed to a poor or very poor median household income change YoY**. To state this again, but in the contrapositive, there were more cases in which eviction rate *decreased* for poor or very poor changes in median household income as opposed to areas with good or very good changes in median household income.

Conclusion

So where have we landed in answering the questions we set out to investigate and with respect to the hypotheses we offered?

1. Commonly held beliefs regarding the rental market and eviction rates are refuted on the aggregate. Low affordability of home ownership does not consistently translate to low affordability of rent, although it does lead to higher rates of rental. And rents that are more affordable see higher levels of eviction than less affordable rents. As each case is delved into deeper, there was a local aspect to the reasoning.

The area of interest that ought to be investigated further is why affordable rent translates to higher eviction rates.

2. The relationship between changes in eviction rate and changes in cost of rent between 2009 and 2011 aren't as straightforward as we supposed at the outset. We busted our hypothesis that the correlation between these two variables was positive (increases in cost of rent are associated with increases in eviction rate and vice versa). We observed, on the contrary, that increases in cost of rent just as often resulted in decreases in rate of eviction as they did increases. When we adjoined the median household income data, we noticed that changes in eviction rate were actually positively correlated with changes in median household income (positive changes in median household income were associated more often with cases of eviction rate increasing and negative changes in median household income were associated more often with cases of eviction rates decreasing).

   If we could investigate further, we'd test the hypothesis that the explanation for the relationship between these variables is that increases in income drive increases in rental demand, which in turn, drives increases in rental rate. Areas with increased eviction rate are more often places where income has increased substantially because landlords can afford to evict non-paying tenants assuming they're more likely to fill their unit given the increase in rent. On the other hand, if income in the area is decreasing, demand is lower, and landlords are more likely to have a vacant unit if they evict their tenants, which is a worse outcome for the landlord than a keeping tenants who pay partially or pay late.