An Exploratory Analysis of Housing Affordability Measures and Homelessness in the US

By Megan Ma

Outline

- Literature Review
- Research Goals / Introduction
- Data Sources
- Visualizations and Summary Statistics
- Linear Regression Model Results
- Additional Figures

Literature Review

How to House the Homeless (Brendan O'Flaherty, 2010) (PDF)

Linear correlation between:

- state median rent and homelessness rate
- median rent to income ratio and homelessness rate

Zillow Research Group (Chris Glynn, Thomas H. Byrne, and Dennis P. Culhane)

- Rent costs are rising but overall homelessness counts are decreasing
- Communities where people spend more than 32 percent of their income on rent can expect a more rapid increase in homelessness

Intro

Research Goals:

- Provided updated statistics/visualizations for 2020 homelessness counts
- Focus on various housing cost and affordability measures outside of just rent and rent to income ratio (e.g. home prices, condo prices) to determine which ones are most correlated to rates of homelessness

About Homelessness Rate Data

HUD Exchange (Source)

- Point-in-Time Estimates by State (2007-2021)

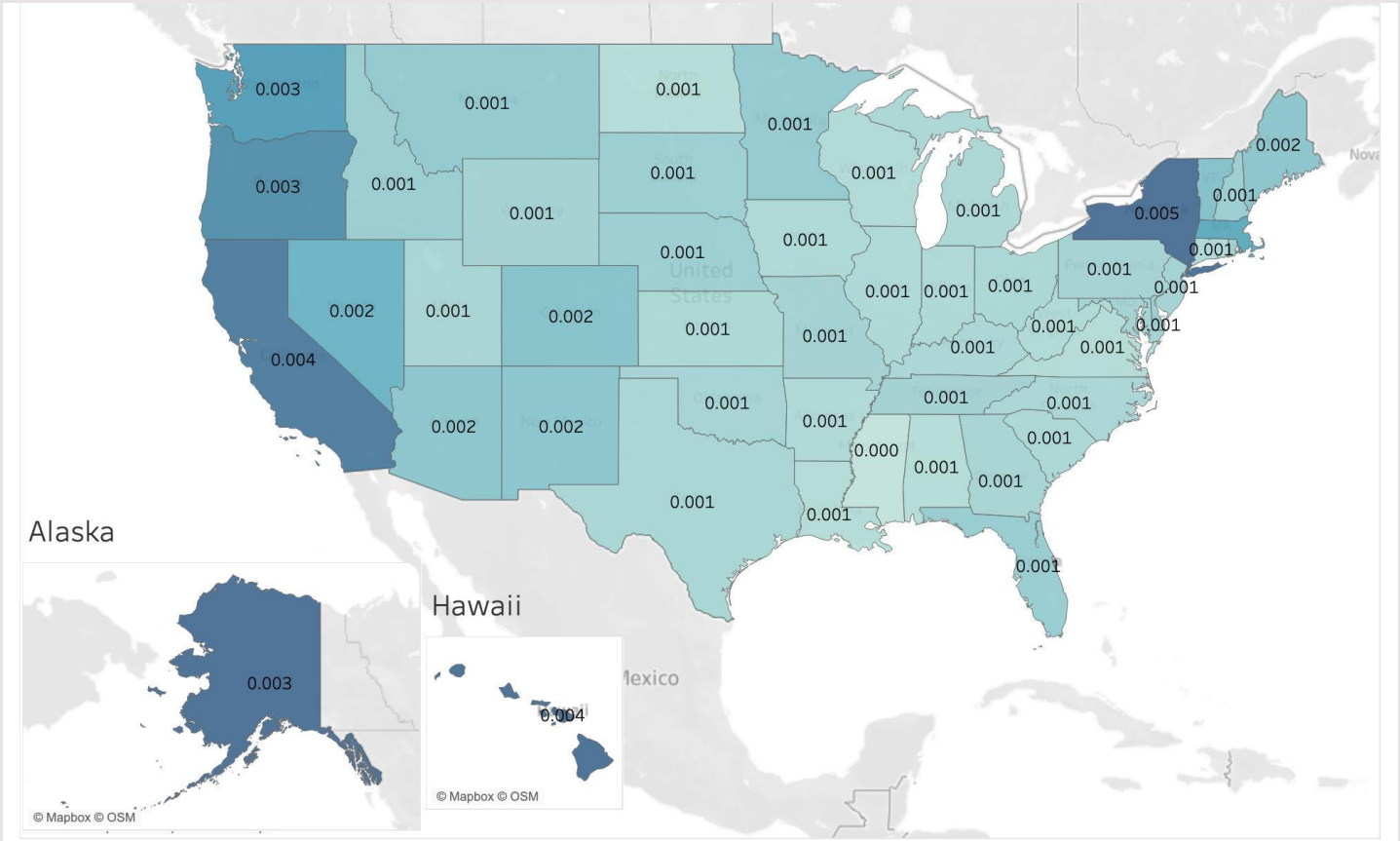
Census Population Estimates (Source)

- Annual Estimates of the Resident population for the United States

$$\text{Homelessness Rate (\%)*} = \frac{\text{PIT Estimates}}{\text{Census State Population Estimate}} \times 100$$

*For each state in 2020

Homelessness Rate (2020)



About Housing Data

Zillow Home Value Index (ZHVI): Source
A smoothed, seasonally adjusted measure of the typical home value across a given region and housing type.

Additional Information: overview of ZHVI and a deep-dive into its methodology.

- Bottom-tier ZHVI: reflects the typical value for homes in the 5th to 35th percentile range.
- Mid-tier ZHVI: reflects the typical value for homes in the 35th to 65th percentile range.
- Top-tier ZHVI: reflects the typical value for homes in the 65th to 95th percentile range.
- Condo/coops ZHVI: typical value for all condo/coops in a given region

American Communities Survey 2020 (ACS)

- Median Rent

About Additional Data

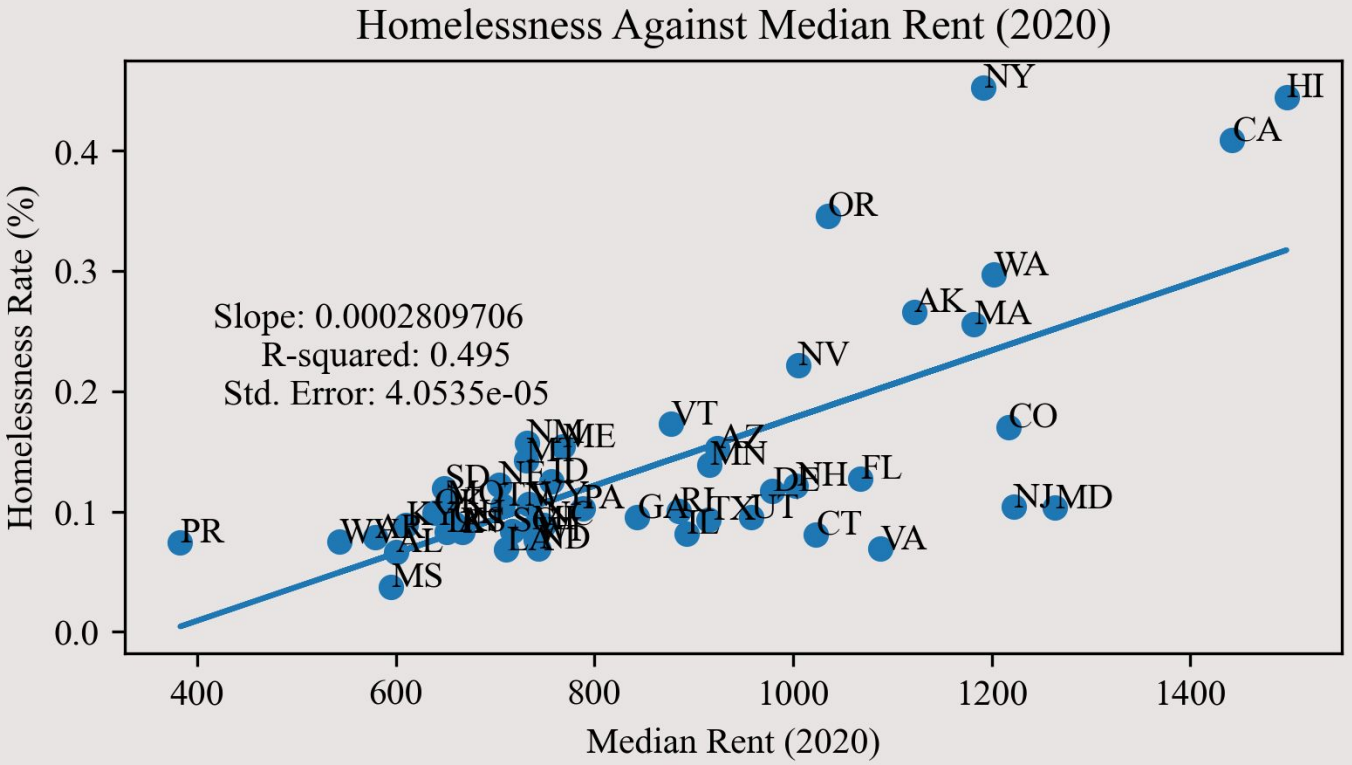
American Communities Survey 2020 (ACS)

- percentage of renters
- percentage of college grads
- percentage of people with incomes at half or less than half of the poverty threshold
- unemployment percentage
- median household income
- Gini coefficient of income inequality
- population density
- percentage of crowded households (units occupied by 1.01 persons or more per room)
- percentage of households with children)

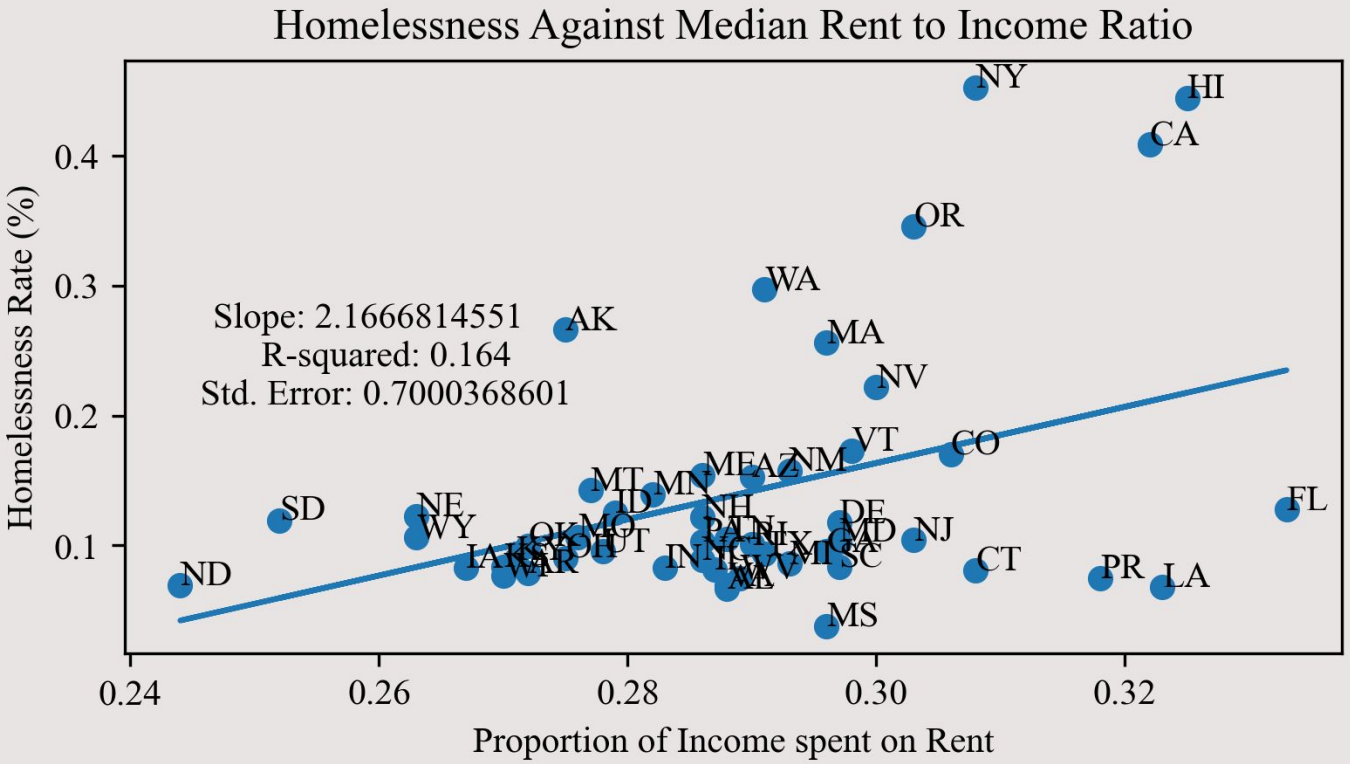
Political Party Data

- Governor political party 2020, 0 for dem 1 for republican
- Percent of 2020 senate members from each state that are democratic
- Percent of 2020 house of representative members from each state that are democratic

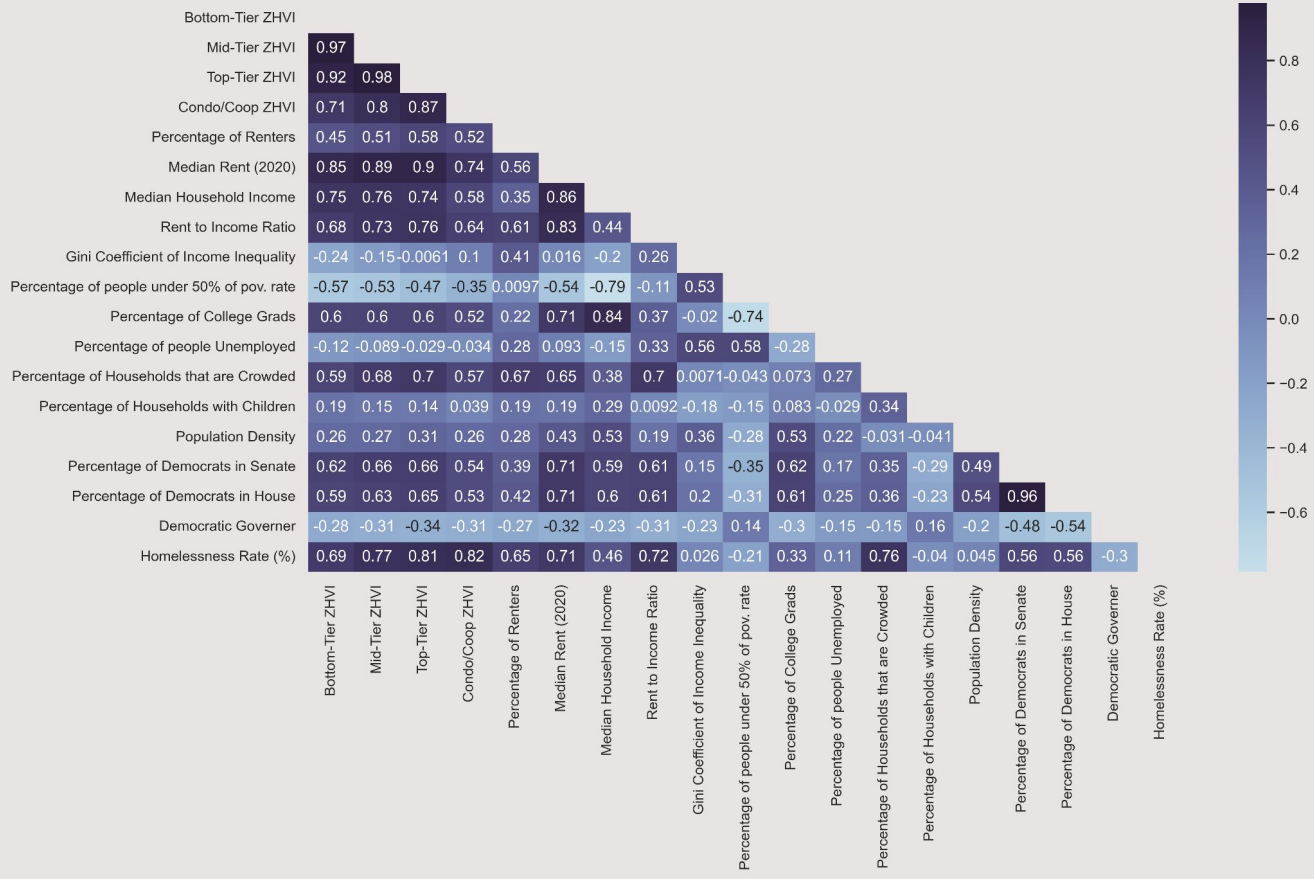
Correlation with Rent



Correlation with Rent to Income Ratio



Correlation Matrix for all Variables



Correlation with Homelessness Rate

	Homelessness Rate (%)
Condo/Coop ZHVI	0.821
Top-Tier ZHVI	0.81
Mid-Tier ZHVI	0.775
Percentage of Households that are Crowded	0.759
Median Rent (2020)	0.71
Bottom-Tier ZHVI	0.69
Percentage of Renters	0.646
Percentage of Democrats in Senate	0.561
Percentage of Democrats in House	0.557
Median Household Income	0.46
Percentage of College Grads	0.327
Percentage of people Unemployed	0.11
Population Density	0.045
Gini Coefficient of Income Inequality	0.026
Percentage of Households with Children	-0.04
Percentage of people under 50% of pov. rate	-0.212
Democratic Governor	-0.297

- High correlation between housing measures and homelessness rate compared to other variables
- Condo/coop and top tier housing have highest correlation with homelessness

Reasonable to expect higher correlation between homelessness and bottom tier/rent prices, so what is going on here?

Possible Explanation: Gentrification

An increase in wealthier, usually white, people arriving in an existing urban neighborhoods causes low income residents to be priced out of their housing, which explains the correlation between high cost housing and homelessness.

Indicators of Gentrification

- Increase in rent/home prices
- Increase in median income rates
- Changes in the makeup of the resident population (race/education level)

→Fit a multiple regression model to get a better handle on what may be going on with those measures with unexpected correlations (Eg. Condo/coop ZHVI)

Model Results

Condo/Coop vs natural log of Homelessness Rate (%)

OLS Regression Results

Dep. Variable:	Ln Homelessness Rate (%)	R-squared:	0.576
Model:	OLS	Adj. R-squared:	0.567
Method:	Least Squares	F-statistic:	63.88
Date:	Tue, 16 Aug 2022	Prob (F-statistic):	2.61e-10
Time:	22:08:20	Log-Likelihood:	23.400
No. Observations:	49	AIC:	-42.80
Df Residuals:	47	BIC:	-39.02
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.3219	0.054	-24.260	0.000	-1.432	-1.212
Condo/Coop ZHVI (in hundred thousands)	0.1746	0.022	7.993	0.000	0.131	0.219

Omnibus:	4.130	Durbin-Watson:	2.288
Prob(Omnibus):	0.127	Jarque-Bera (JB):	3.009
Skew:	0.488	Prob(JB):	0.222
Kurtosis:	3.721	Cond. No.	7.06

A $100,000 increase in Condo/Coop Price results in a 0.1746% increase in homelessness rate

Model Results

Condo/Coop vs natural log of Homelessness Rate (%) with gentrification variables

OLS Regression Results

Dep. Variable:	Ln Homelessness Rate (%)	R-squared:	0.661
Model:	OLS	Adj. R-squared:	0.630
Method:	Least Squares	F-statistic:	21.47
Date:	Wed, 17 Aug 2022	Prob (F-statistic):	7.08e-10
Time:	10:36:51	Log-Likelihood:	28.889
No. Observations:	49	AIC:	-47.78
Df Residuals:	44	BIC:	-38.32
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.6129	0.265	-6.091	0.000	-2.147	-1.079
Median Rent (2020)	0.0002	0.000	0.809	0.423	-0.000	0.001
Rent to Income Ratio	29.3726	22.323	1.316	0.195	-15.616	74.361
Condo/Coop ZHVI (in hundred thousands)	0.1153	0.030	3.850	0.000	0.055	0.176
Percentage of College Grads	-0.0047	0.007	-0.705	0.485	-0.018	0.009

Omnibus:	1.176	Durbin-Watson:	2.421
Prob(Omnibus):	0.556	Jarque-Bera (JB):	0.489
Skew:	-0.165	Prob(JB):	0.783
Kurtosis:	3.361	Cond. No.	1.00e+06

Adding variables related to gentrification lowers the effect of Condo/Coop Prices on homelessness rate

Some, but not all, of the correlation between Condo/Coop Prices can be explained by these variables. There is still some unexplained correlation that should be investigated further.

Model Results

Condo/Coop vs natural log of Homelessness Rate (%) with gentrification variables + additional variables

OLS Regression Results

Dep. Variable:	Ln Homelessness Rate (%)	R-squared:	0.727
Model:	OLS	Adj. R-squared:	0.672
Method:	Least Squares	F-statistic:	13.29
Date:	Wed, 17 Aug 2022	Prob (F-statistic):	3.90e-09
Time:	10:37:51	Log-Likelihood:	34.144
No. Observations:	49	AIC:	-50.29
Df Residuals:	40	BIC:	-33.26
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.4524	0.333	-4.367	0.000	-2.125	-0.780
Median Rent (2020)	-0.0007	0.000	-1.638	0.109	-0.001	0.000
Rent to Income Ratio	65.5010	26.148	2.505	0.016	12.654	118.348
Percentage of College Grads	0.0059	0.009	0.637	0.528	-0.013	0.025
Condo/Coop ZHVI (in hundred thousands)	0.1024	0.031	3.346	0.002	0.041	0.164
Percentage of Households that are Crowded	6.1797	2.520	2.452	0.019	1.086	11.273
Percentage of people under 50% of pov. rate	-0.0784	0.040	-1.954	0.058	-0.160	0.003
Percentage of people Unemployed	0.0173	0.033	0.530	0.599	-0.049	0.083
Democratic Governer	-0.0062	0.043	-0.145	0.885	-0.092	0.080

Omnibus:	0.128	Durbin-Watson:	2.180
Prob(Omnibus):	0.938	Jarque-Bera (JB):	0.074
Skew:	0.080	Prob(JB):	0.964
Kurtosis:	2.897	Cond. No.	1.24e+06

Adding variables that are not necessarily related to gentrification lowers the coefficient for condo/coop variable even more

There may be a more complex mechanism behind the correlation between condo/coop prices and homelessness rate than just gentrification

Future Research

Additional Qualitative research to develop a theory for explaining the link between high cost housing and homelessness (outside of gentrification)

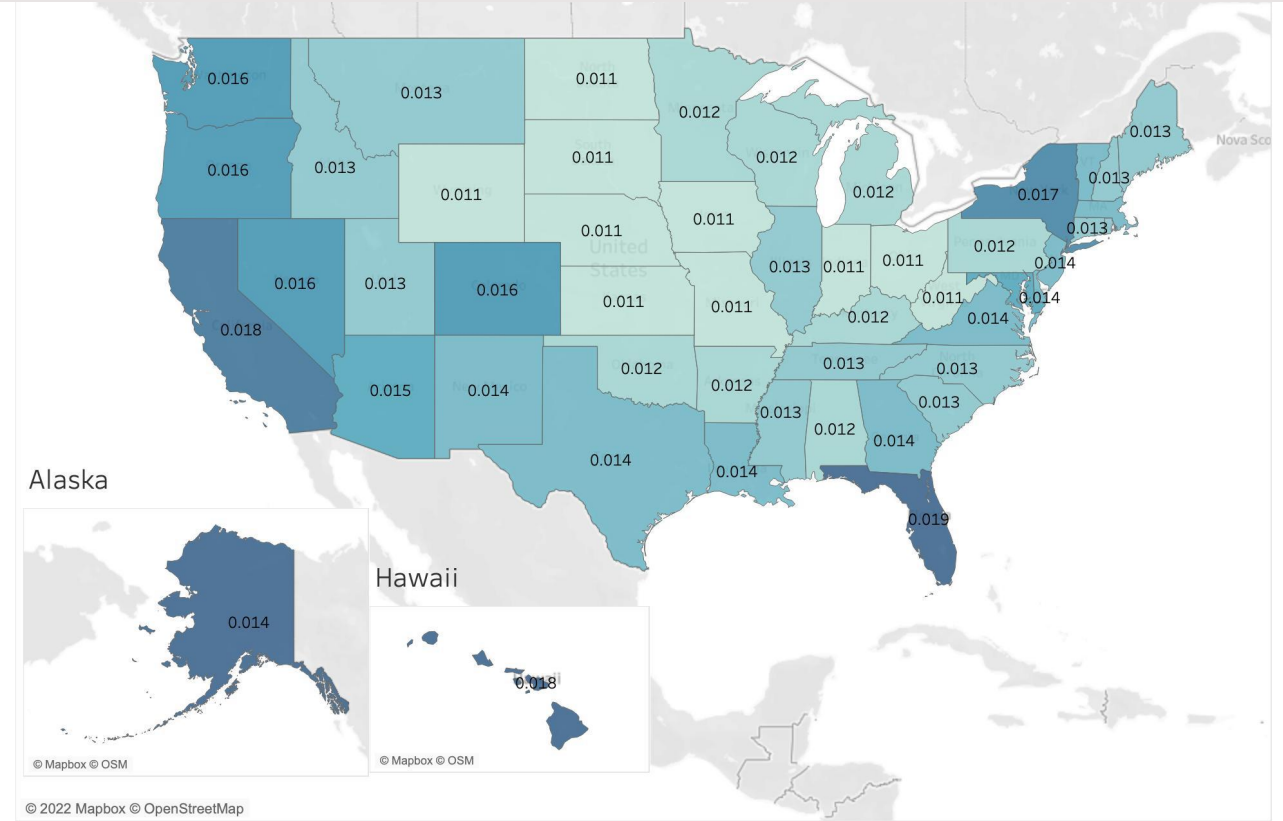
Clustering by smaller geographic locations

Model for systemic undercount of the PIT homelessness count

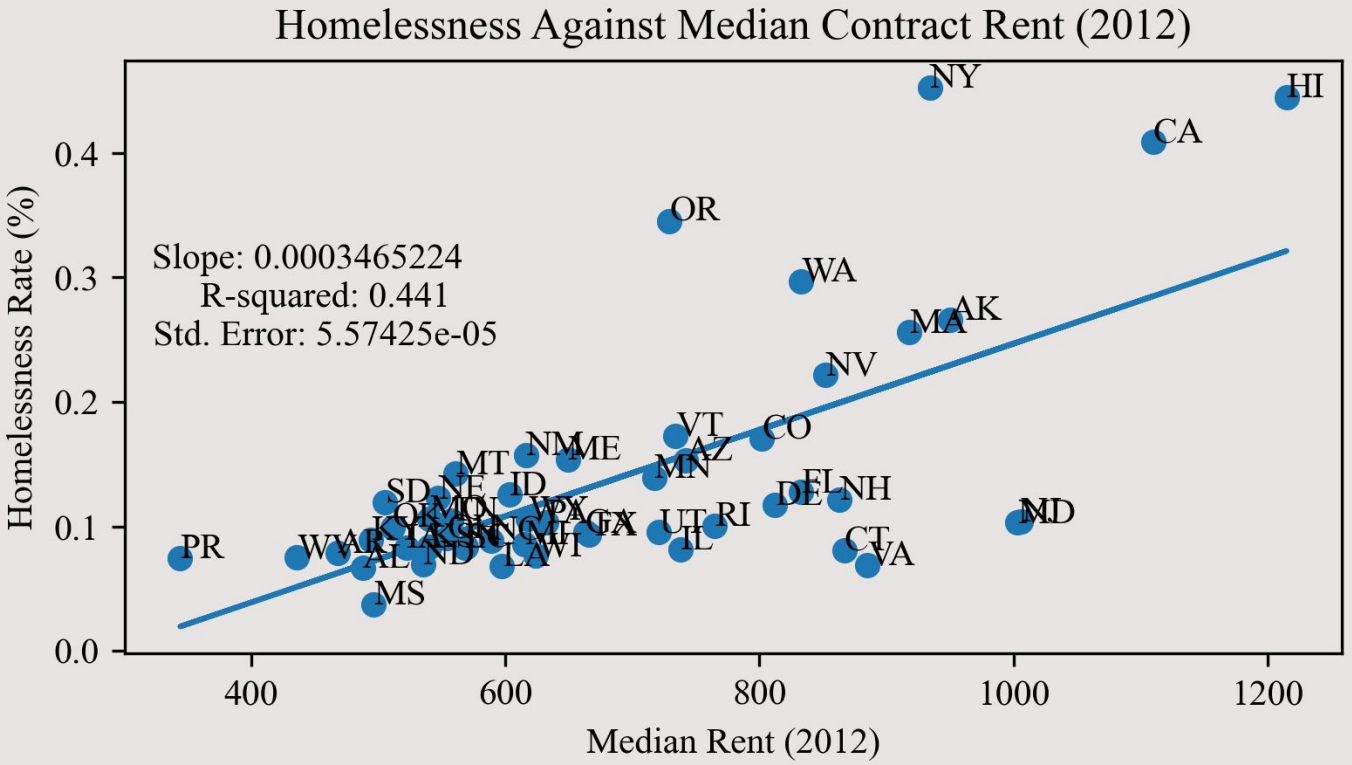
Test various models which might be better suited than linear regression (eg. Random forests, neural networks → finding correlations between variables)

Fixed effects for changes over time

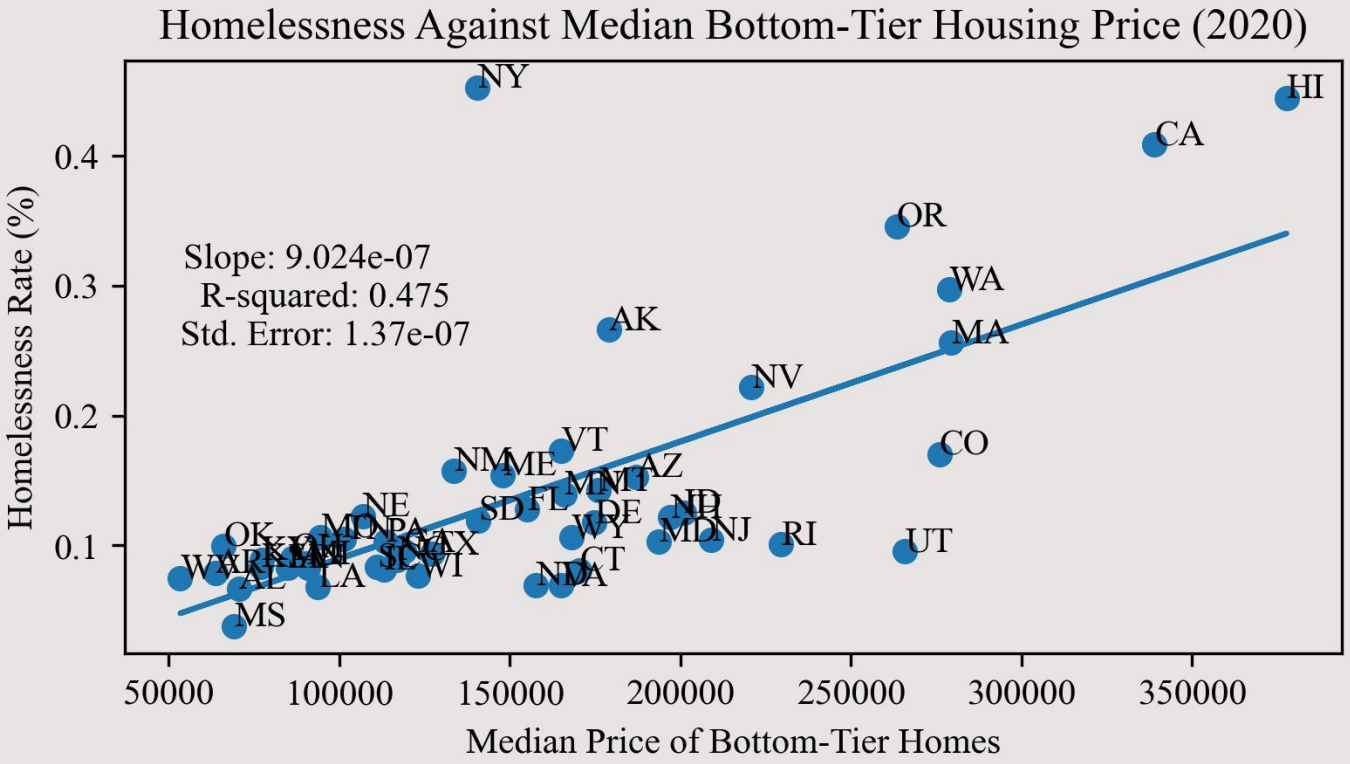
Additional Figures: Rent to Income Ratio



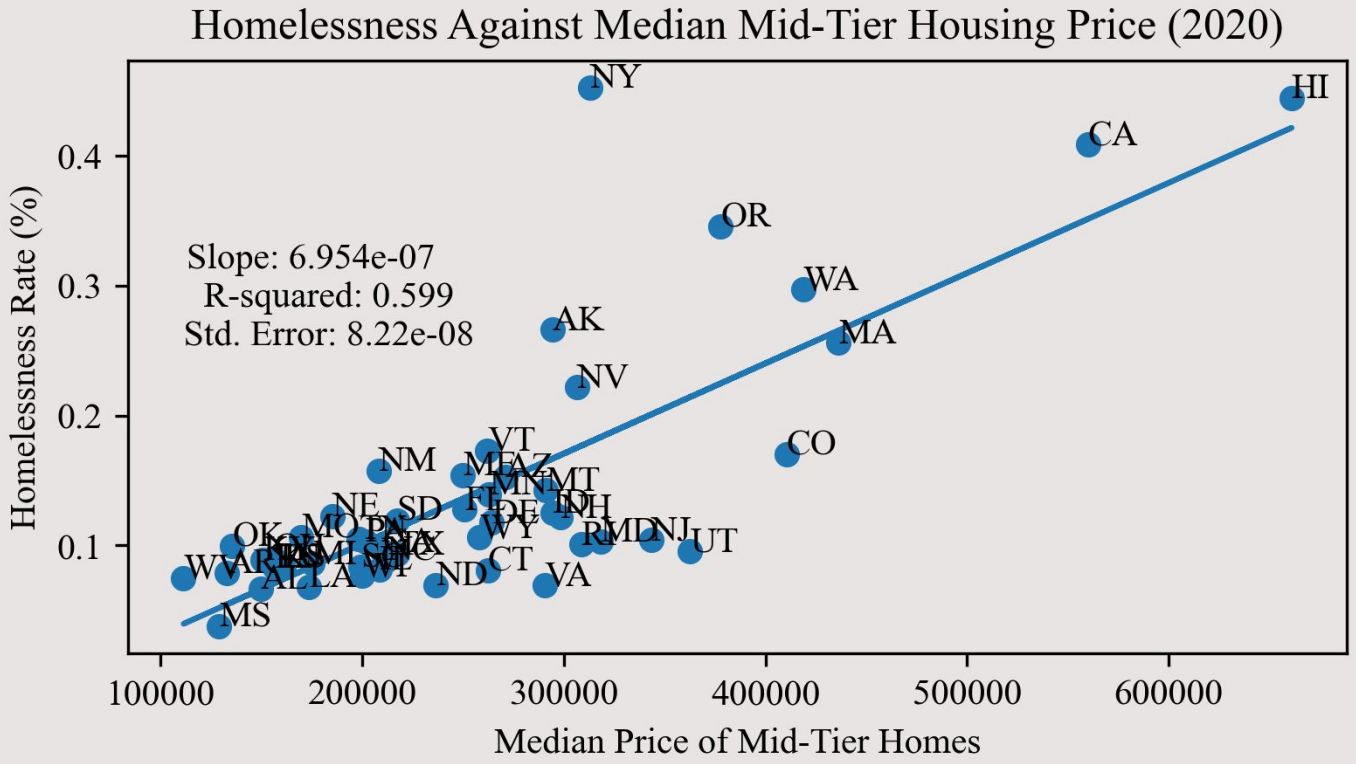
Additional Figures: Correlation with Median Rent (2012)



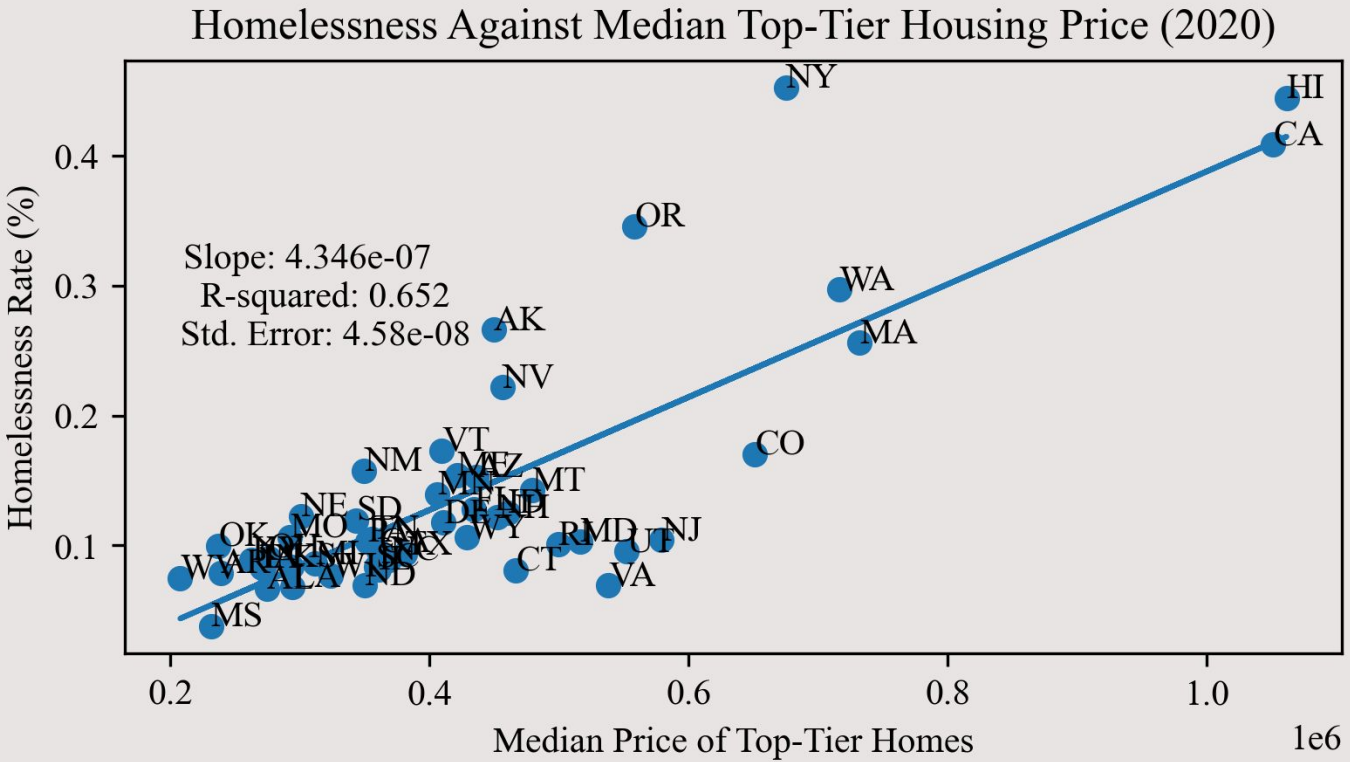
Additional Figures: Correlation with Bottom-Tier ZHVI



Additional Figures: Correlation with Mid-Tier ZHVI



Additional Figure: Correlation with Top-Tier ZHVI



Additional Figures: Correlation with Single Family ZHVI