

Capstone Project Phase 3

Kimberly Najarian

2022-04-28

Introduction

This project focuses on homelessness and see how it correlates to housing prices to see if the rising price in houses is causing more people to be homeless. To do this, I will find a data set on the number of homeless people per state and another data set on the housing price data per state. With this information, I will see if a solution to homelessness is by having affordable housing available. I hypothesize that with the data I will accumulate, there will be a correlation between the housing prices and correlation. I have heard a theory about this correlation but I have not seen any studies about it.

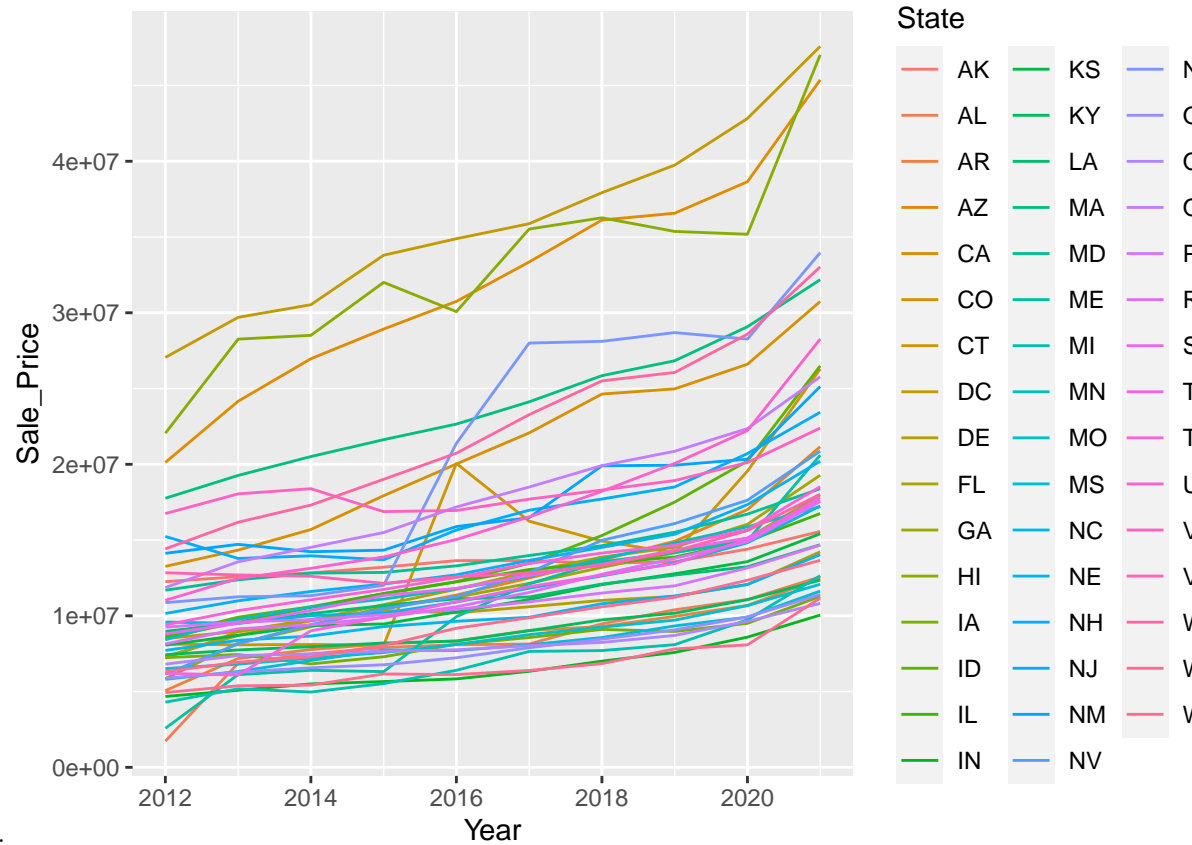
Data

I used the housing data from Redfin and the Homeless data from a website called the HUD Exchange which helps provide housing for the homeless. To see if the trends of both the homeless and housing data correlate with the GDP, I also pulled in a data set of GDP by state from the Bureau of Economic Analysis website and formatted the data in Excel to correlate with the other data sets. I used the Total Sheltered Homeless variable for the homeless data, the Median Sale Price for the housing data, and the GDP for the GDP data. I compared all of this with State and Year. The following is the summary statistics for the data.

##	Data	mean	sd	n	max	min	median
## 1	totalHomeless	8.762671e+02	4753.798	423	76051	0.0	36
## 2	SalePrice	1.455038e+07	7589563.145	423	47584600	4970500.0	12576500
## 3	GDP	3.759381e+05	456331.320	423	2871424	28648.5	233016

There are some limitations to the data. The homeless data only counts sheltered homeless and does not include people living on the streets or in tents, which seems to be a large number of people. Another limitation is that there may be other causes for the correlation that is not included in the data I have. ## Empirical Results

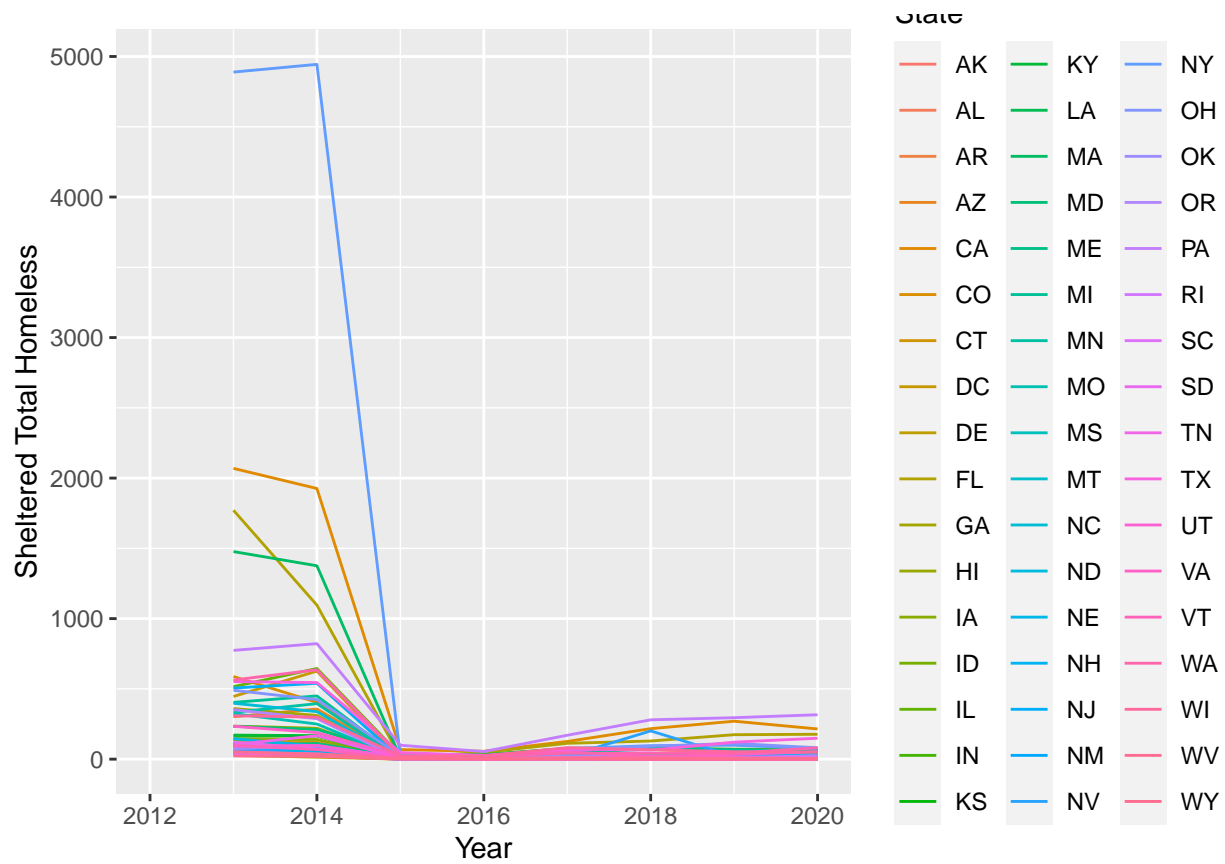
I used some graphs to help visualize the data to see if there are any correlation in the trends of homelessness and



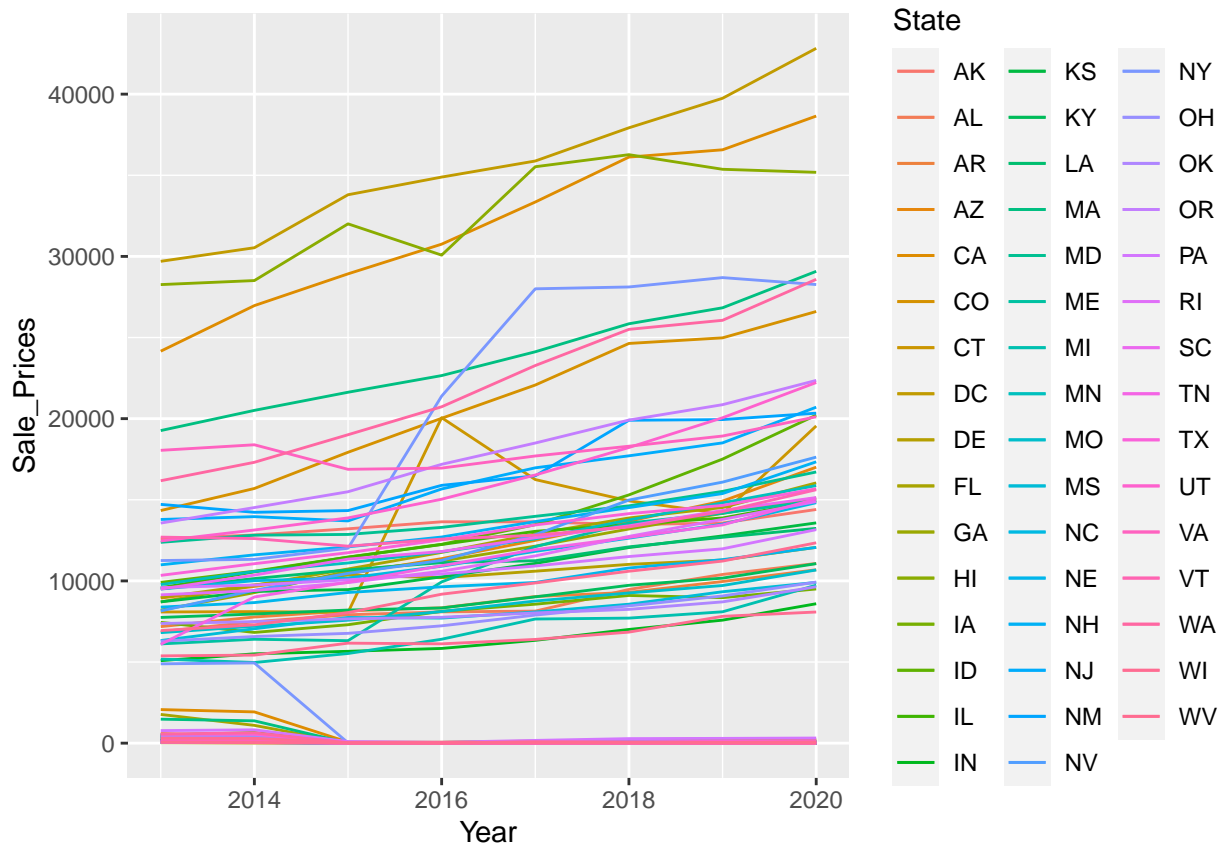
housing over the years.

The first graph is of the housing data with each line representing a state.

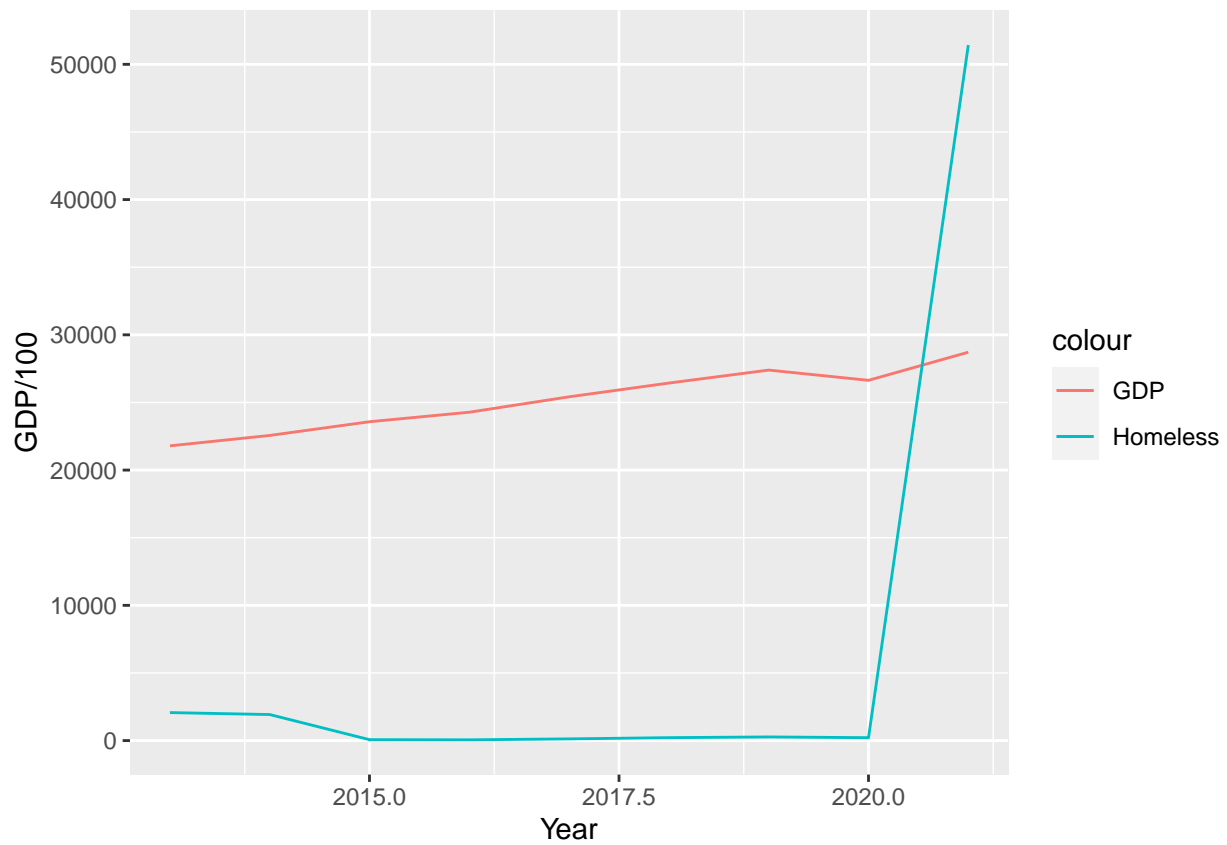
Warning: Removed 51 row(s) containing missing values (geom_path).



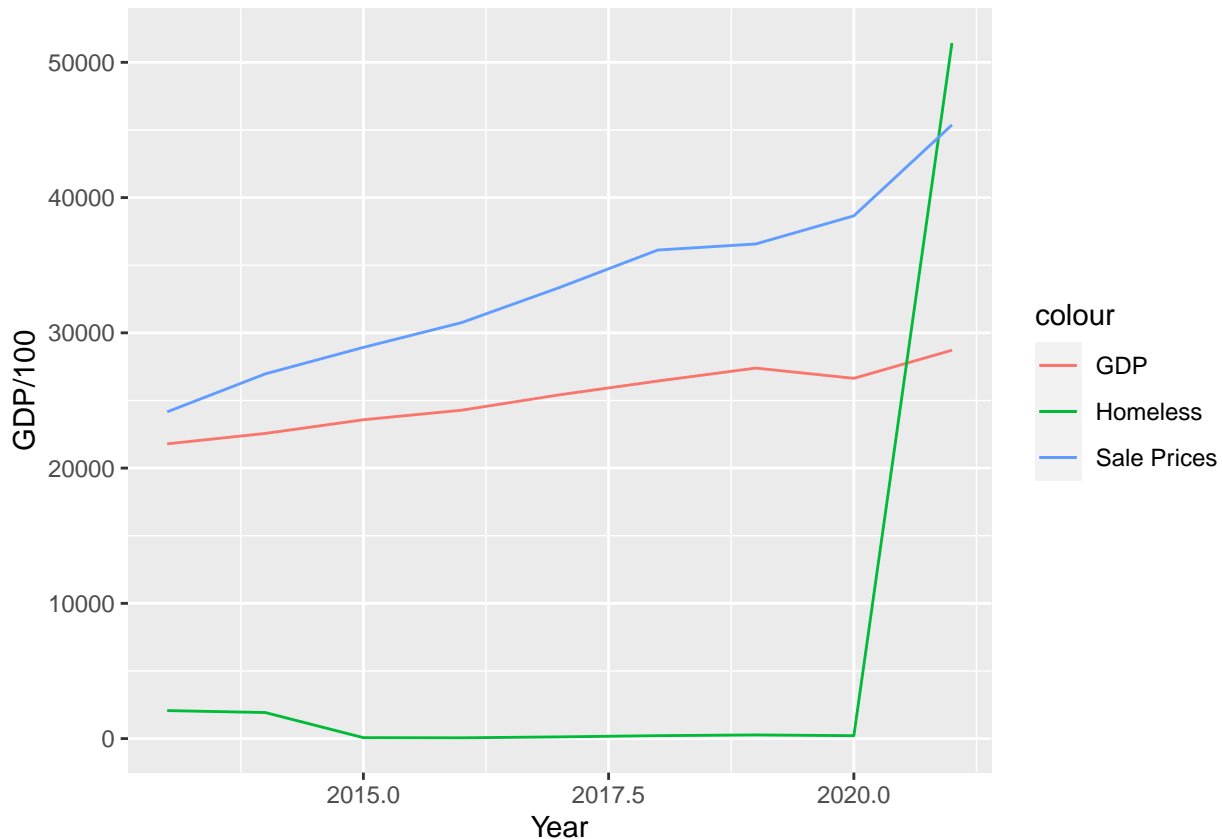
This graph is a plot of the homeless population per state and year. It seems like a lot of people were able to leave the homeless shelters in 2015 but the populations started rising again in 2017.



Here is a graph of each state with the housing prices and homeless populations combined. There doesn't seem to be much correlation between the housing prices and homelessness but there does seem to be areas where line slopes change at the same time. This is why I added in the GDP data to see if they correlate with the GDP changes per state which I will show later on.



Here is a graph of the home prices and homeless population in just California. I included 2021 here. I didn't before because of the drastic increase that year which made the data between 2015 and 2020 look flat and at 0. There might be a slight correlation, but doesn't appear to be much.



Here is a graph of the California data with GDP included. Both the homeless and housing prices seem to somewhat correlate with the GDP.

```
##
## Call:
## lm(formula = `Sheltered Total Homeless.y` ~ GDP + Sale_Price,
##     data = HHG)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##    -9140    -936     -53     424    70019
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.926e+03  4.718e+02  -4.083 5.33e-05 ***
## GDP          2.305e-03  5.020e-04   4.591 5.85e-06 ***
## Sale_Price   1.331e-04  3.019e-05   4.408 1.33e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4460 on 420 degrees of freedom
## Multiple R-squared:  0.1241, Adjusted R-squared:  0.1199
## F-statistic: 29.75 on 2 and 420 DF, p-value: 8.249e-13
```

According to the linear model, there is a significant correlation between GDP, House prices, and Homelessness. The R squared is pretty small but not too much to mean they aren't at least somewhat correlated.

```
##
## Call:
```

```
## lm(formula = GDP ~ `Sheltered Total Homeless.y` + Sale_Price,
##     data = HHG)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -839766 -225545  -90488   60801 2034047
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.318e+05  4.517e+04   2.918  0.00371 **
## `Sheltered Total Homeless.y` 2.073e+01  4.516e+00   4.591 5.85e-06 ***
## Sale_Price      1.553e-02  2.829e-03   5.490 6.98e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 423000 on 420 degrees of freedom
## Multiple R-squared:  0.1449, Adjusted R-squared:  0.1408
## F-statistic: 35.59 on 2 and 420 DF,  p-value: 5.267e-15
```

I wondered if the homeless and housing correlated more with GDP rather than GDP and housing correlating with homelessness, so I conducted another linear model. Here it does seem that both homeless and housing are dependent on the GDP as the correlation of them both is stronger. Since the R squared is at 0.14, there seems to be other factors that contribute to the causation of the difference but that is beyond the scope of this study.

In all, I believe that my hypothesis was not completely proven right. I found that the homeless population ebbs and flows with the economy, at least for those people who are in the homeless shelters. As it appeared in the data, the Pandemic caused a lot more people to be homeless but did not affect the housing prices at all, so homelessness is more correlated with the GDP than the housing prices.

Conclusion

From this study, I analyzed the correlation between housing prices, GDP, and homeless populations. I found that there is a significant correlation between the three but they mainly depend on the GDP which is the economy. Homelessness especially correlates with GDP. From this result, we can learn that more people become homeless when the economy is bad. With this knowledge, we can try to help those who are struggling with money when we see signs that the economy is suffering.

Future study can include finding data on the unsheltered homeless population to see what those number trends look like, and with that, look at data of drugs or things like that to see if they correlate. As of now, the main takeaway is that people really depend on the economy to be able to afford housing.