Gentrification Within the U.S.

Meiling Traeger & Damian Franco

Outline

- Introduction
- Project Overview
- Methods
- Findings
- Conclusion

Gentrification

Gentrification can be characterized as the process of high-income groups moving into low-income urban areas.

This often results in existing residents, usually minority groups, becoming displaced and shifting the economic landscape of the area, i.e. more expensive businesses being established.

The United States have been susceptible to gentrification. In fact, 2021 has seen the greatest increase in housing prices over the past year within the U.S.. Although this is partially due to the pandemic, the rising cost of living is not a recent trend.

Gentrification

Characteristics of Gentrification
- Increased investment in neighborhood amenities, like transit and parks
- Industrial land could change to restaurants and storefronts
- Investors flipping properties for large profits
- High-end development which leads to landlords looking for higher-paying tenants
- High median household cost

Core causes for Gentrification
- Higher income residents
- Rent/Wage gap
- Real estate investments
- Higher income residents moving in
- Pricier business locations
- Education level
- Racial demographic
- Historic conditions (policies and practices that could make communities susceptible)
- Rising median household cost

Project Overview

We plan to implement information theory components and analyze data to help us understand more about areas of gentrification and the factors that contribute to gentrification within the United States. Numerical values will help other, and ourselves, understand that gentrification can be possibly and reasonably quantified.

Finding a correlation between various characteristics of an area and gentrification is our main goal. This will give us insight on what characteristics appear to be more involved in gentrification and could possibly provide others with insight on whether the area they live in is gentrified or could be gentrified.

Human population dynamics are very complicated and there are seemingly endless factors that could contribute to the gentrification of an area. Complex systems, such as the one we are investigating, are very difficult to condense to a small set of characteristics, but we still want to attempt to do so because gentrification is a important topic within the United States.

Quantification

Quantifying gentrification is difficult. We settled on a model that would ensure that we can make a fair assumption to some of the most important factors that contribute to gentrification.

Our area will be *eligible* for gentrification if the population is greater than 500, the median home value is less than the 40% percentile of the United States that year and the median household income is also less than the 40% percentile. Education will not be accounted for when considering eligibility

Areas will be considered *gentrified* is the population is above 500, the median home value and education (bachelor's degree of higher) stats are above the 60% percentile of the United States, and the median household income has increased from the previous year.

Requirements

	Eligible	Gentrified
Population	> 500	
Median Home Value (Percentile)	< 40%	> 60%
Median Household Income (Percentile)	< 40%	Increased
Education (Percentile)		> 60%

Data Sets Used

All datasets provided by the US Census
- Population/Racial demographic
- Education Data
- Income/Employment data
- Household Pricing data

Data sets were chosen to account for the top characteristics that we considered the most important contributing factors to gentrification. The area size that we settled on was county size. Every county in the United States will be analyzed. Originally we wanted to have a smaller size that could contribute to more accurate areas of gentrification, but was unable to properly implement this change with the time given. Whole counties are rarely considered gentrified, but there is still some great inferences and observations that came out of using county size areas.

Shannon Entropy (Methods)

The measure of "uncertainty"

Shannon Entropy, H, measures basic
information capacity.
- For a discrete random variable X with a
  probability mass function p(x)
- Entropy is measured in bits
- H measures the average uncertainty in the
  random variable.

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Transfer Entropy (Methods)

$$T_{X \to Y} = H\left(Y_t \mid Y_{t-1:t-L}\right) - H\left(Y_t \mid Y_{t-1:t-L}, X_{t-1:t-L}\right)$$

- Transfer entropy from a process X to another process Y is the amount of uncertainty reduced in future values of Y by knowing the past values of X given past values of Y.

- Transfer entropy is conditional mutual information where the history of Y (the influenced variable) is in the condition.

H(X)   H(Y)

H(X|Y)   I(X;Y)   H(Y|X)

H(X,Y)
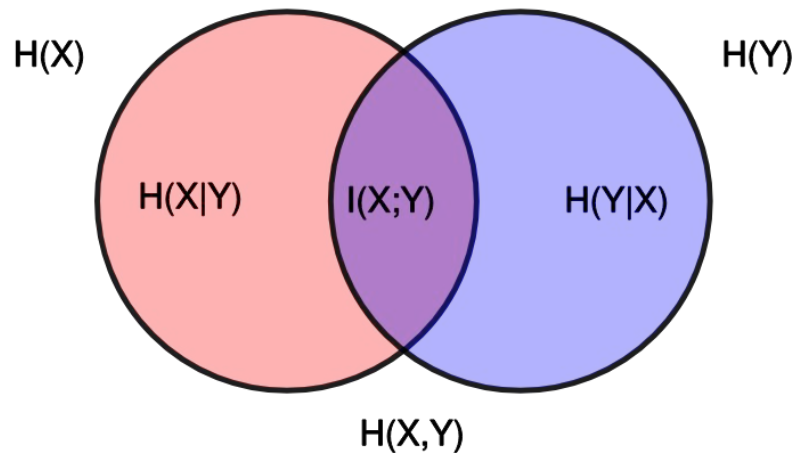
$$T_{X \to Y} = I(Y_t; X_{t-1:t-L} \mid Y_{t-1:t-L})$$

Mutual Information (Methods)

$$I(X;Y) \equiv \mathrm{H}(X) - \mathrm{H}(X|Y)$$
$$\equiv \mathrm{H}(Y) - \mathrm{H}(Y|X)$$
$$\equiv \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y)$$

- Joint: H(X,Y): area in both circles (all colors)
- Entropy of X: H(X), red circle (red + purple)
- Entropy of Y: H(Y), blue circle (blue + purple)
- Conditional H(X|Y): red crescent
- Conditional H(Y|X): blue crescent
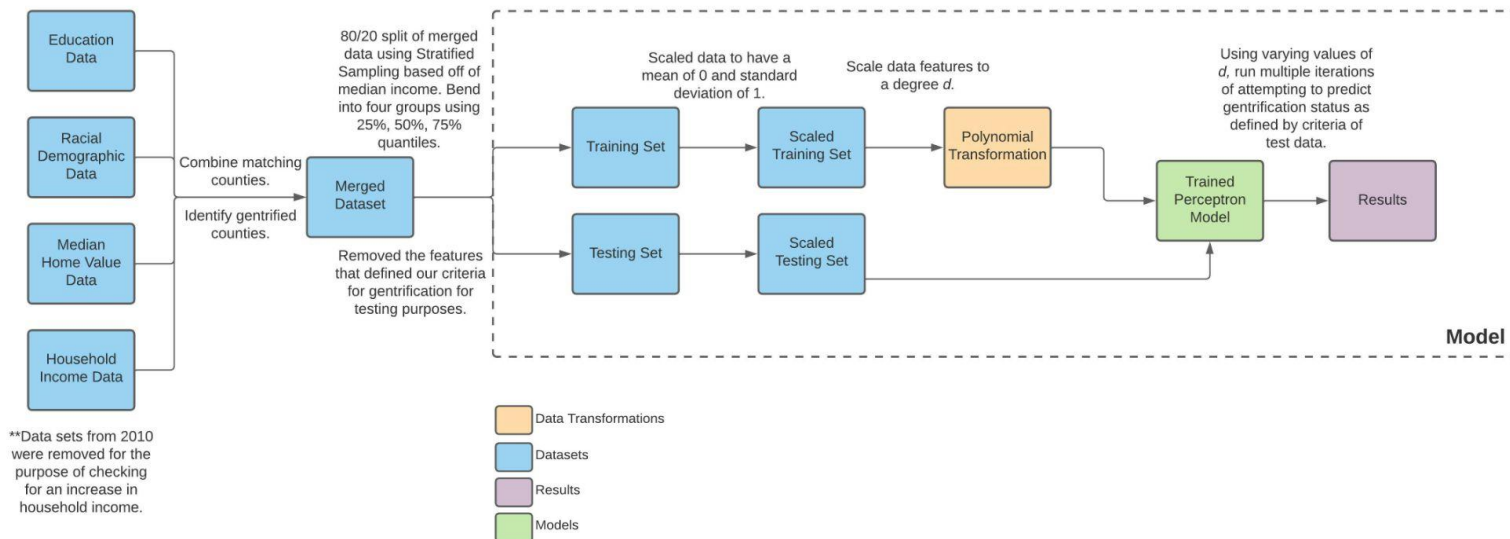- Mutual Information I(X;Y): purple intersection
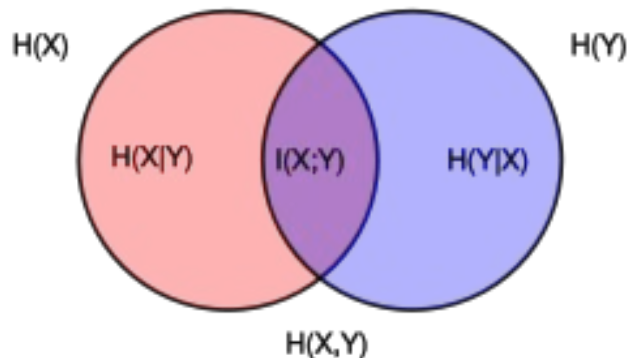
Polynomial Regression (Methods)

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_N x^N + \varepsilon$$

Data model below is our model that will help us understand how we can possibly predict gentrified areas with the use of polynomial regression. This approach uses polynomial regression and a single perceptron model.

Findings

Gentrified, eligible or neither trends (All US Counties)

First we wanted to get a full understanding of the scope of the project. Through calculations of percentiles, we were able to find where all counties lie regarding the median income and education data. We choose to focus on these two factors and leave out median household price because our requirements section requires that these two factors must be above a certain percentile to either be eligible or gentrified. You can see that the density falls below the 60th percentile of both education and household income, but outliers to appear to exist within this data.

Plot above shows all trends for all gentrified, eligible for gentrification or neither counties. Convergence here is something interesting to note. All counties were placed in a category based on the requirements table shown earlier in the presentation.
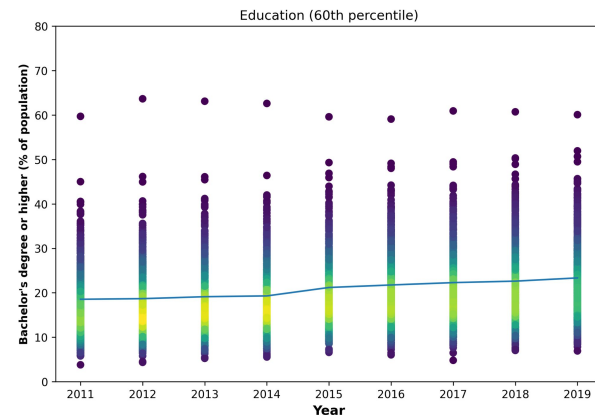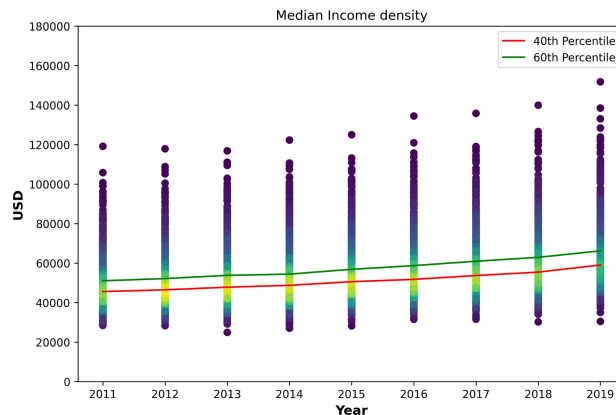
Sample Case

Within this sample case we look into four distinct counties. Two of these counties are considered gentrified (Los Angeles & Denver) and the other two are considered non-gentrified (Milwaukee & Bernalillo). Many of you might be able to make an assumption on which counties are considered gentrified. Those areas are Los Angeles and Denver counties. As you might be able to see, both of the "gentrified" counties have a much higher median home value and median household income than the other two "non-gentrified" counties. Education trends between these counties do differ with Denver being an outlier.
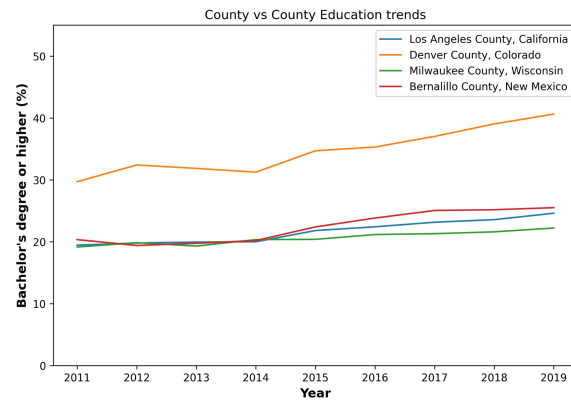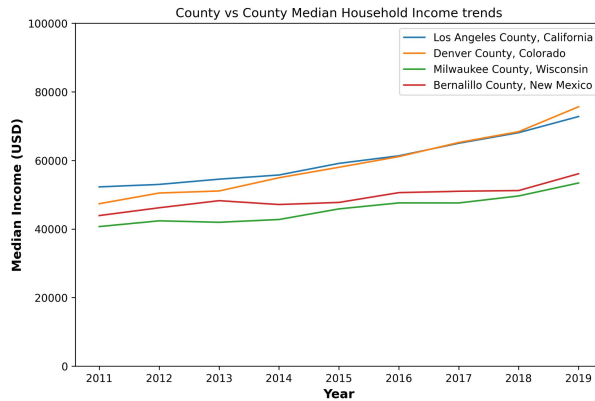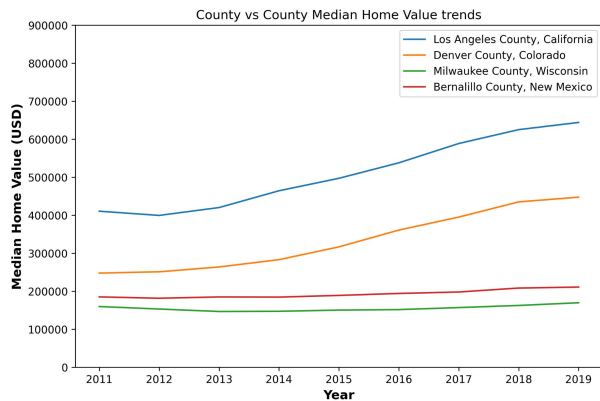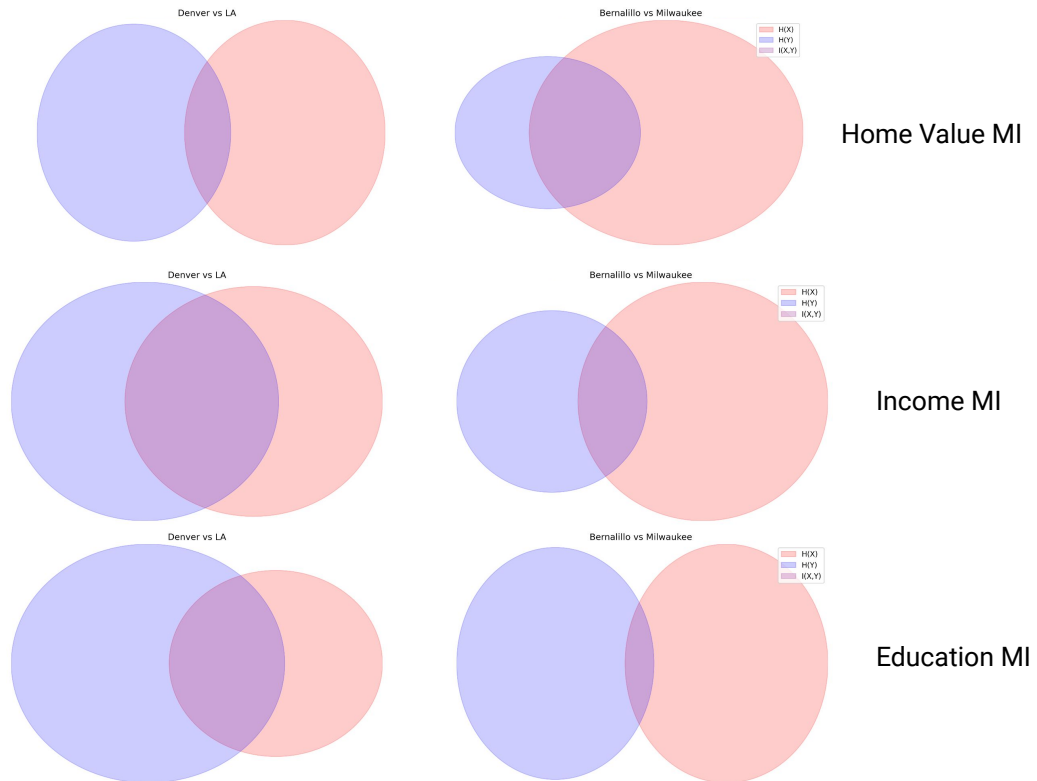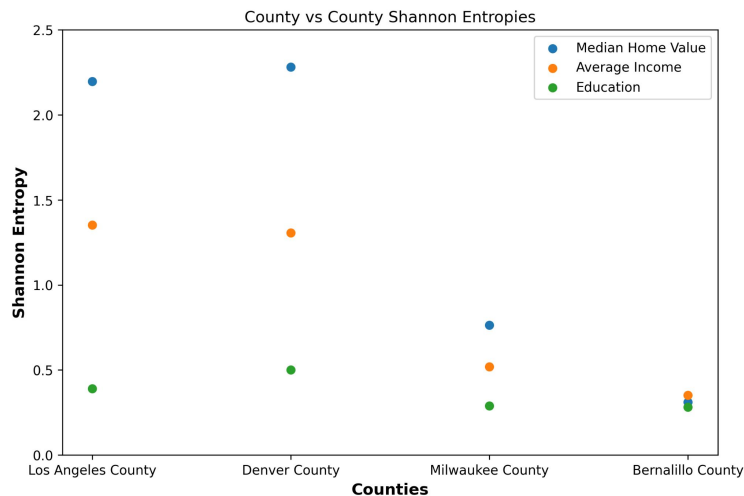
Sample Case Cont.

County Names:
Los Angeles County, California
Denver County, Colorado
Milwaukee County, Wisconsin
Bernalillo County, New Mexico

Eligiblity?
[0, 0, 0, 0, 0, 0, 0, 0, 0]
[0, 0, 0, 0, 0, 0, 0, 0, 0]
[0, 0, 0, 0, 1, 1, 1, 1, 1]
[0, 0, 0, 0, 0, 0, 0, 0, 0]

Gentrified?
[0, 1, 1, 1, 1, 1, 1, 1, 1]
[1, 1, 1, 1, 1, 1, 1, 1, 1]
[0, 0, 0, 0, 0, 0, 0, 0, 0]
[0, 0, 0, 0, 0, 0, 0, 0, 0]

Findings

The heat map below shows every single counties transfer entropies for average home value, average household income and average education. Dark hexagons on the heat plots indicate high concentrations of values within that area.
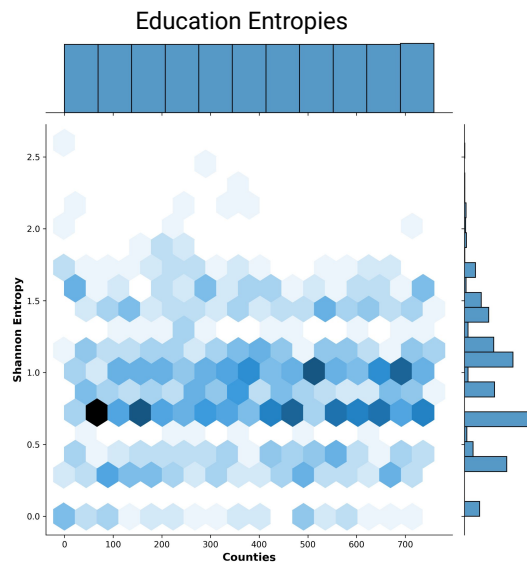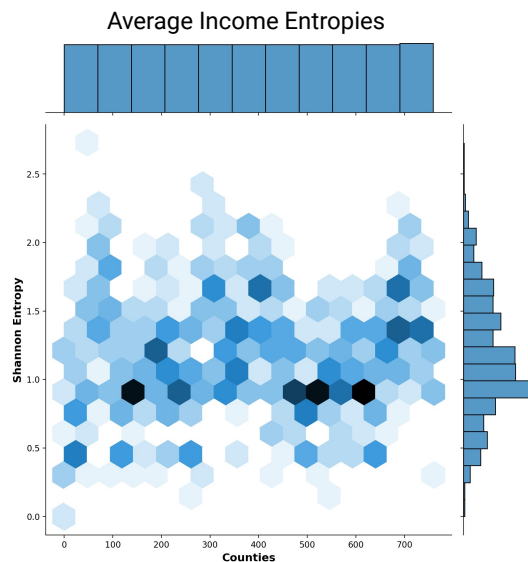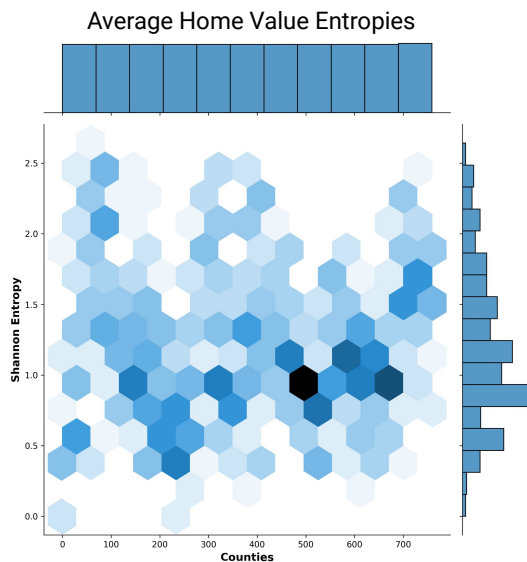
Findings

Degree (d)	Testing Score
1	.6724
2	.699
3	.6573
4	.7175
5	.6752
6	.6903
7	.6609
8	.6581
9	.6767
10	.6344

Predictions that we have been forecasting has shown some promise, but not very high promise. We believe that the outliers may have to do with the lower testing scores within our polynomial regression model. Polynomial regression has known to be faulty when outliers appear within the data set that it is trying to train. We are assuming that this is where our problem lies, but we are still proud of the somewhat accurate gentrification status prediction.

Conclusions

- Human population dynamics are complicated.
- Areas of gentrification exist in smaller areas then counties and county sized data is too large of an area to make an assumption.
- Entropy has proven to be useful when investigating why some areas are gentrified.
- The dynamics of education, median household income and average home value could possibly indicate that an area can become gentrified if implemented well.
- Importance of racial demographic is still unknown.
- Both information theory components and polynomial regression could be combined to more accurately predict gentrification status.

Thank you for listening!
Any questions?