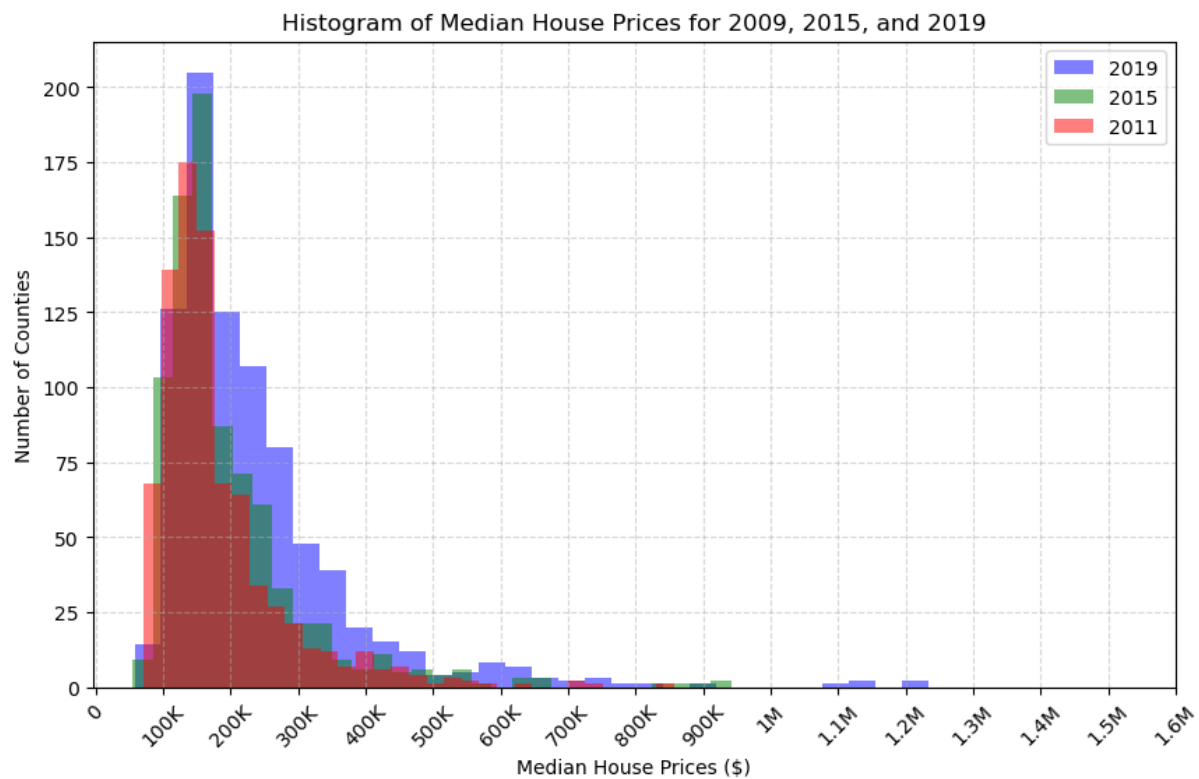


Question 1 – How have median house prices and our variables changed over time?

Figure 1

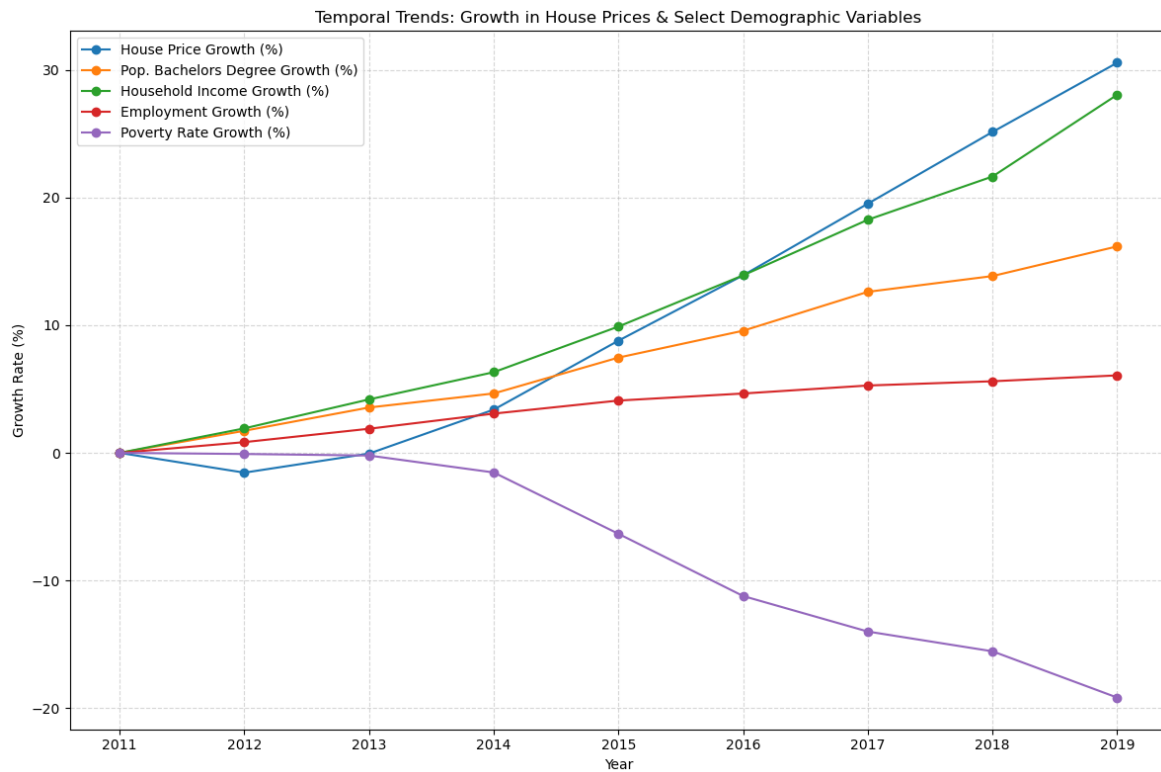


The histogram displays house prices across U.S. counties for three different years: 2011 (red), 2015 (green), and 2019 (purple). These are the initial, mid-point, and final years in our data, respectively. The x-axis shows median home prices across counties range from \$0 to \$1.6 million.

Trends Observed

1. 2011: Most counties had median house prices between \$100k and \$200k.
2. 2015: There is a noticeable shift with fewer counties having median house prices below \$100k and an increase in the number of counties with house prices in the \$200k to \$300k range, indicating an overall increase in house prices.
3. 2019: The distribution shifts further, with most counties having median house prices between \$200k and \$300k (more breadth in prices). There's also a noticeable increase in the number of counties with median house prices above \$300k, reflecting a trend of rising house prices over the years.

Figure 2



This line graph illustrates the cumulative growth rates of five key variables relative to the base year 2011, spanning to 2019. The variables analyzed are:

1. House Price Growth (%) (Blue)
2. Population with Bachelor's Degree Growth (%) (Orange)
3. Household Income Growth (%) (Green)
4. Employment Growth (%) (Red)
5. Poverty Rate Growth (%) (Purple)

Observations

1. House Price Growth (%): There is a steady and substantial increase post 2012 in house prices over the years, reaching the highest growth rate by 2019. The rate of increase is constant from 2015 and onwards.
2. Population with Bachelor's Degree Growth (%): This variable shows a consistent upward trend, though at a slower rate compared to house price growth. The steady increase suggests a gradual improvement in educational attainment levels across the USA over the years.

3. Household Income Growth (%): Like the house prices, household income growth shows a steady increase over time. The positive trend indicates rising incomes, potentially contributing to higher affordability and thus higher house prices. This may suggest this variable is more closely related to house prices.
4. Employment Growth (%): Employment growth exhibits a moderate increase over the years, reflecting overall economic recovery and job creation. Employment levels can directly influence household incomes and, consequently, housing demand and prices.
5. Poverty Rate Growth (%): Interestingly, this variable shows a significant decline over the years. The reduction in poverty rates suggests an improvement in economic conditions, which may correlate with rising household incomes and employment, and therefore house prices.

Implications

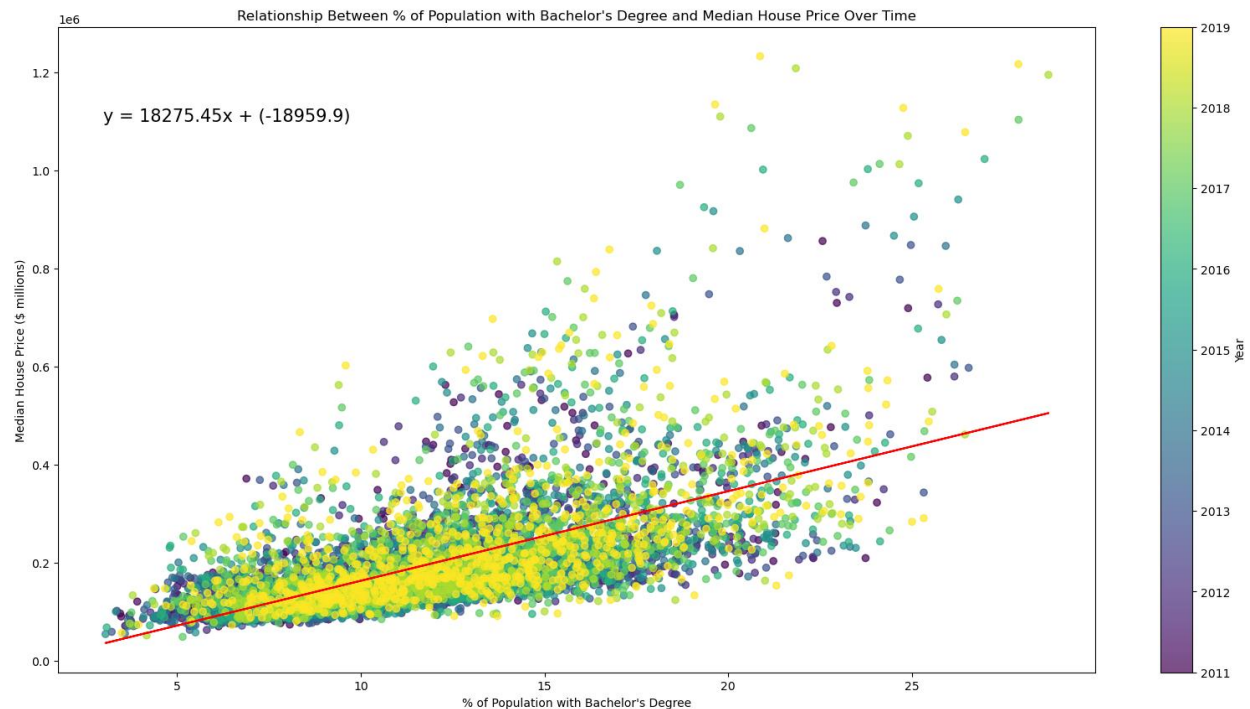
- The steady increase in house prices, income, education levels, and employment suggests improving economic conditions, particularly in the post-GFC era. These factors likely interplay, driving demand for housing and contributing to price appreciation.
- The reduction in poverty rates paired with rising educational attainment points to broader social improvements. As more people attain higher education, their earning potential increases, which could lead to a more stable economic environment.

Conclusion:

Over the period from 2011 to 2019, there is a clear trend of increasing median house prices. In addition, the distribution of house prices has shifted to higher price brackets over time. We can also see that educational attainment, income and employment have moved upwards over time in tandem with median housing prices. These improvements in economic fundamentals are also evident via the reduction in poverty rates over time.

Question 2 - How do our selected demographic variables correlate with median house prices across counties in the USA?

Figure 1 - Education and House Prices



This scatterplot plots median house prices across counties in the USA for every year between 2011 and 2019. Every year has a color according to the gradient scale on the right.

The slope of 18275.45 suggests that, on average, for every 1% increase in the population with a bachelor's degree, the median house price increases by \$18,275.45

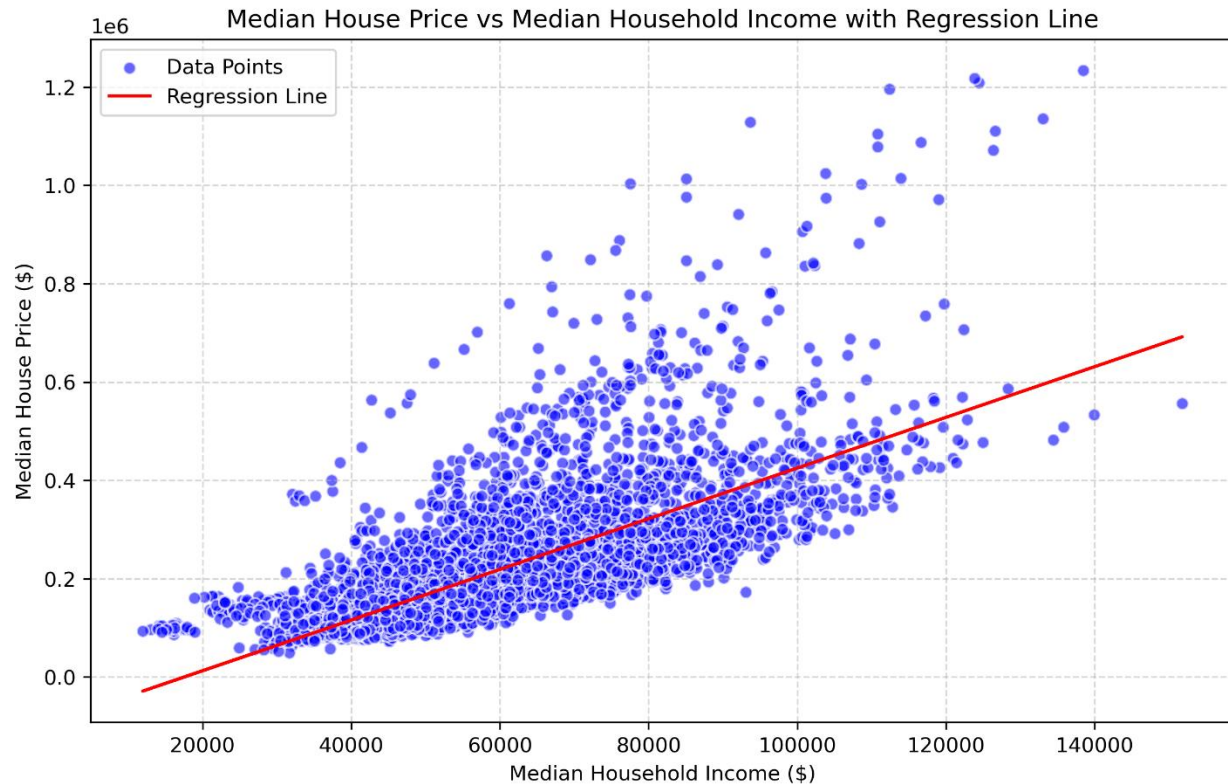
The correlation between these variables is 0.63 and the R-squared is 0.3998. The p-value of the correlation coefficient is 0.000.

The negative intercept of -18959.9 suggests that this analysis is not quite valid for small house prices and in counties with near-zero bachelor's degree holder's rates.

There are some visible outliers at the higher price points (above ~\$700k price points). We believe leaving these in the analysis is informative given they provide insight into certain exclusive markets, which are part of the social fabric.

We conclude that the relationship between both variables is medium in strength, and that at least at third of the variability in prices can be captured by educational attainment

Figure 2 – Median Income and House Prices



The scatter plot shows a positive relationship between Median Household Income and Median House Prices. Areas with higher household incomes generally tend to have higher house prices, which aligns with the expectation that wealthier areas can sustain higher housing market values.

The points are spread upwards as income increases, indicating a positive correlation. There may be some areas where house prices deviate from the general trend, suggesting other influencing factors like location desirability or market conditions.

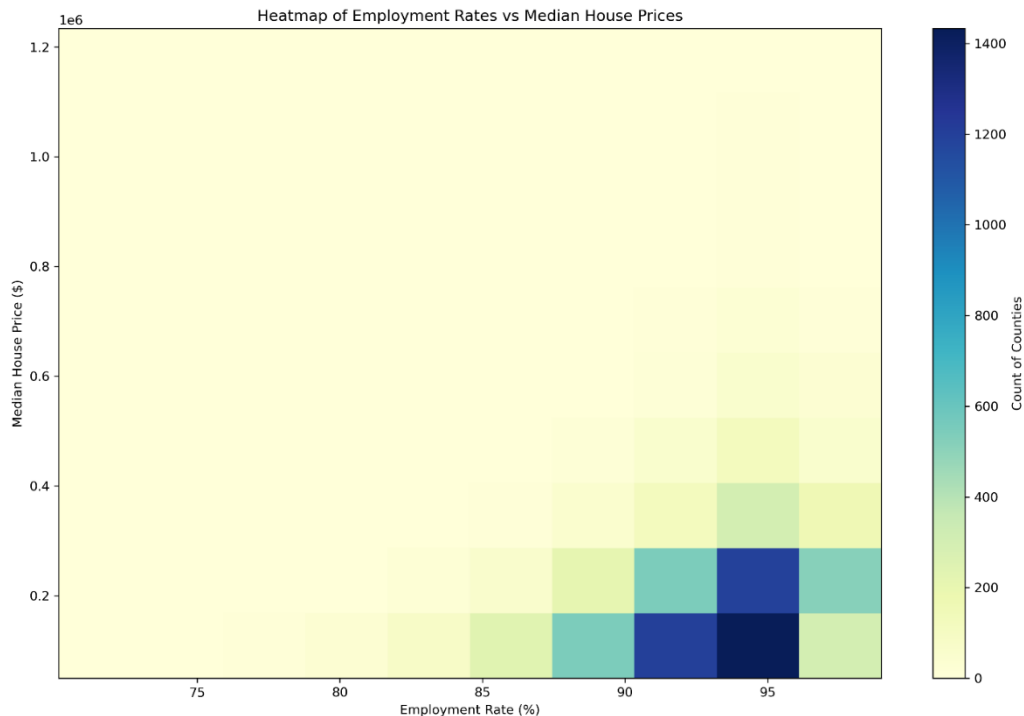
The red regression line represents the linear relationship between income and house prices. The data points generally align with the line, indicating a positive correlation where house prices increase as household income rises. The R^2 value (indicated in the legend) quantifies how well the line explains the variability in house prices based on income.

Slope (5.15): For every \$1 increase in median household income, the median house price increases by approximately \$5.15 on average.

The correlation coefficient (R) of 0.71 indicates a strong positive linear relationship between median household income and house prices.

The R-squared value of 0.51 implies that 51% of the variability in house prices is explained by the median household income. While significant, there may be other factors influencing house prices.

When the median household income is \$0 (theoretical scenario), the model predicts the median house price would be approximately -\$90,380.97. This value isn't realistic but is common in regression when the data



doesn't cover the lower range of the independent variable.

Figure 3 – Employment Rates and House Prices

There is a weak positive correlation (0.26) between employment rates and house prices, suggesting that regions with higher employment rates tend to have slightly higher house prices. However, the relationship is not strong, indicating that other factors, such as demand driven by external investments or local economic conditions, likely play a more significant role.

The heatmap shows that regions with the highest employment rates (above 90%) tend to have clusters of higher house prices. However, the increase in house prices with the employment rate is not linear.

Most of the density is concentrated in mid to high employment rate ranges (85–95%), where house prices remain moderate. Over time, areas with higher employment rates may attract more residents, driving up housing demand and prices. However, areas with extreme employment rates (very low or very high) show mixed patterns, as other economic or geographic factors likely dominate.

The heatmap indicates relatively few regions with both low employment rates (below 80%) and high house prices. These regions may exist but are outliers, as most high house prices are clustered with high employment rates.

Figure 4 – Poverty Rates



Regression - Statistical Results

- r-value (Correlation Coefficient): -0.44
- r^2 -value (R-squared): 0.19
- p-value: 0.0000

The scatter plot depicts the relationship between the percentage of people living in poverty (x-axis) and the median house price (y-axis) from 2011 to 2019. The data points are color-coded by year, and a logarithmic regression line in red represents the trend.

Key Results from the Regression:

- R-value (Correlation Coefficient): -0.44
 - Interpretation: The negative correlation coefficient indicates a moderate inverse relationship between poverty rates and median house prices. As the percentage of people living in poverty increases, the median house price tends to decrease.
- R-squared (R^2): 0.19
 - Interpretation: The R^2 value of 0.19 suggests that approximately 19% of the variability in median house prices can be explained by the variability in poverty rates. While this indicates a significant relationship, it also implies that 81% of the variability is due to other factors not included in this model.

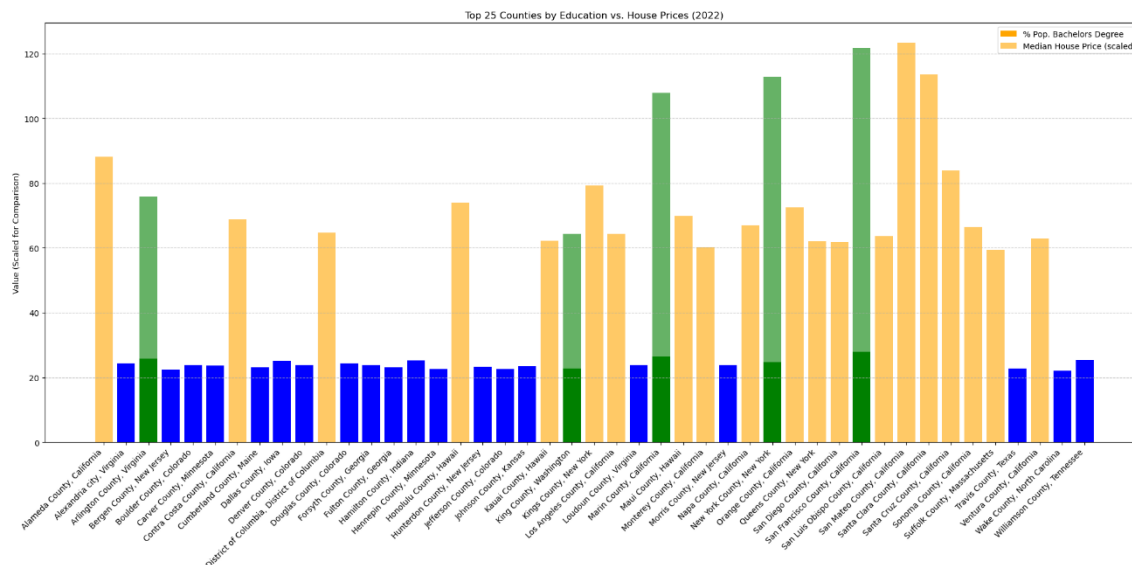
- P-value: 0.0000
 - Interpretation: The p-value of nearly zero means that the relationship between poverty rates and median house prices is statistically significant. This confirms that the observed trend is not due to random chance.

Insights from the Graph:

1. **Inverse Relationship:** The regression line shows a clear inverse relationship. As poverty rates increase, median house prices tend to drop, especially noticeable at lower poverty percentages where house prices decrease rapidly. This trend stabilizes as poverty rates increase further.
2. **Yearly Color Coding:** The color coding by year demonstrates that this relationship holds consistently over time, with slight variations across different years, reflecting broader economic conditions and market dynamics.
3. **Trend Consistency:** The consistency of the trend line across the years suggests that the relationship between poverty and house prices remains relatively stable over the observed period, reinforcing the reliability of the findings.

Implications: High poverty rates correlate with lower median house prices, which could be due to a lower overall demand for housing in impoverished areas, potentially coupled with less investment in real estate development.

Question 3 - Are median house prices higher in counties with a more educated population?



RESULTS OF A T TEST ON THE DIFFERENCE BETWEEN MEAN HOUSE PRICES FOR TOP AND BOTTOM 25% COUNTIES

- T-statistic: 15.76392606893622
- P-value: 2.851220545245644e-44

Analysis of the T-Test Results in Relation to the Bar Chart

The T-test conducted compares the median house prices for counties in the top 25% of educational attainment with those in the bottom 25%. From the results, we can conclude that there is a significant difference in median house prices between top and bottom educational attainment counties.

- The extremely low p-value (much less than 0.05) indicates a highly significant difference between the mean house prices of the top 25% and bottom 25% educational attainment counties. This strong significance suggests that the educational attainment level is indeed a factor affecting house prices.
- The large T-statistic value further supports that the difference in house prices between these groups is material.

Relationship to the Bar Chart

The bar chart shows the distribution of counties based on their house prices and educational attainment:

- **Green Bars** (Counties in both top 25% for education and house prices) make up roughly 20% of the total bars, highlighting an overlap between high education levels and high house prices.
- **Blue Bars** (High education, not high house prices) and **Yellow Bars** (High house prices, not high education) show the presence of counties where the correlation does not align perfectly, indicating other influencing factors.

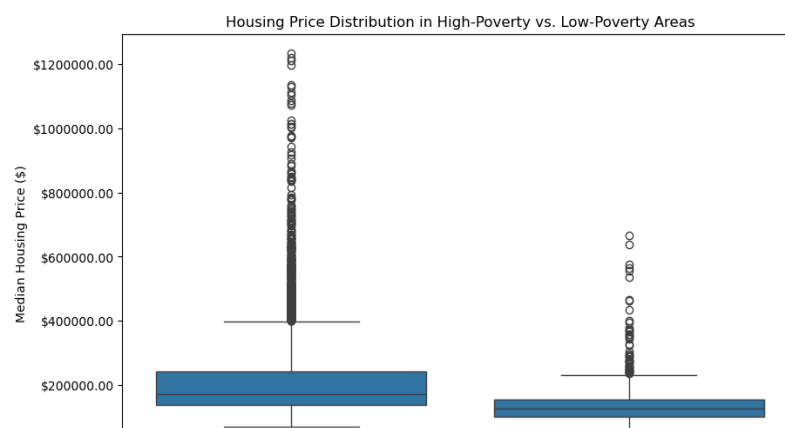
Integrating T-Test Results with Bar Chart Insights

- **Consistent Findings:** The T-test results align with the visual insights from the bar chart. Both the statistical analysis and the chart suggest that higher educational attainment is associated with higher house prices, but it's not the only factor at play.
- **Other Influences:** The presence of blue and yellow bars indicates that while education plays a significant role, other variables also affect house prices, such as local economic conditions, geographical desirability, and employment opportunities.
- **Moderate R-Squared Value:** The R-squared value of 0.3998 from the scatter plot further supports the notion that education explains a significant portion of the variability in house prices but not all of it. This aligns with the roughly 20% green bars, suggesting a moderate correlation.

Conclusion

The combination of the T-test results and the bar chart provides a robust understanding of the relationship between education levels and house prices. The significant difference in house prices between high and low educational attainment counties reinforces the importance of education. Still, the presence of counties that do not fit this trend highlights the multifaceted nature of the housing market.

Question 4 - What is the distribution of housing prices in high-poverty vs. low-poverty areas?



The box plots compare the distribution of median house prices in low-poverty areas (left) versus high-poverty areas (right). Here's an expanded analysis based on your initial commentary:

Key Observations

1. Median Housing Prices:

- Low-Poverty Areas (poverty rates under 20%): The median housing price is substantially higher, indicating that areas with lower poverty rates generally have more expensive homes.
- High-Poverty Areas (poverty rates over 20%): The median housing price is lower, which aligns with the expectation that higher poverty rates correlate with lower housing prices.

2. Range and Outliers:

- Low-Poverty Areas: The range of housing prices is broader, signifying greater variability in house prices. Numerous outliers extend up to approximately \$1,200,000, indicating the presence of extremely high-value properties. These outliers highlight significant disparities in property values within low-poverty areas.
- High-Poverty Areas: The range of housing prices is more restricted, with fewer outliers and lower maximum prices. This suggests a more constrained housing market where property values do not reach the high extremes seen in low-poverty areas.

3. Interquartile Range (IQR):

- Low-Poverty Areas: The IQR is wider, suggesting that there is more variability in the middle 50% of house prices. This greater spread indicates diverse housing market conditions in low-poverty areas.
- High-Poverty Areas: The IQR is narrower, indicating less variability within the central range of house prices. This narrower spread implies a more uniform market where house prices are more similar.

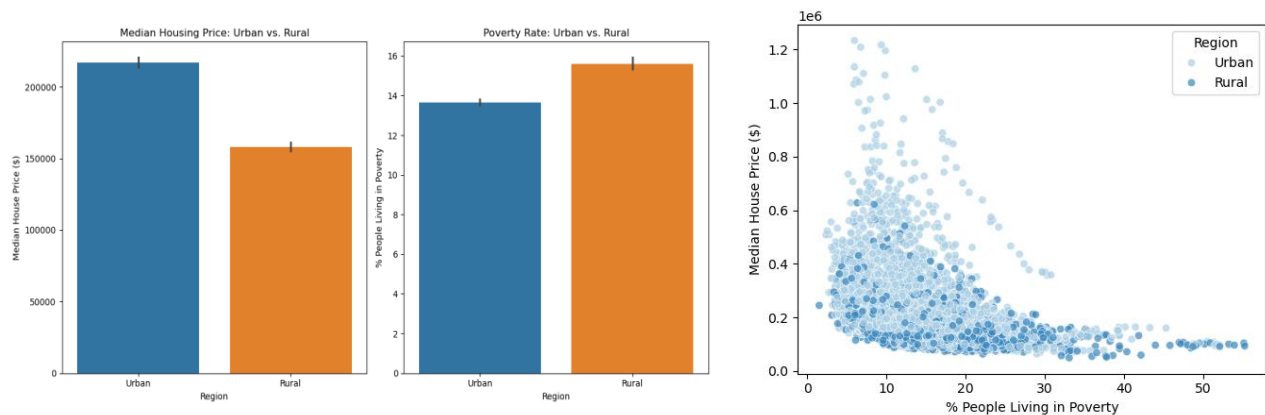
4. Outliers:

- Low-Poverty Areas: Many outliers represent extremely expensive properties. These outliers indicate that certain properties in low-poverty areas significantly exceed the typical market range, potentially driven by luxury housing markets or highly desirable locations.
- High-Poverty Areas: There are fewer outliers, and those present are closer to the central range of housing prices. This suggests that even the highest-priced properties in high-poverty areas are not as far removed from the typical market values, reflecting more constrained housing market dynamics.

Conclusions

- **Economic Disparities:** The box plots visually underscore the economic disparities between low- and high-poverty areas. Higher median housing prices in low-poverty areas can be attributed to higher demand, better amenities, and overall economic prosperity.
- **Market Variability:** Greater variability in house prices in low-poverty areas may reflect a diverse range of property types, from affordable homes to luxury estates, while high-poverty areas tend to have more homogeneous housing markets.

Question 5 - How do median house prices and poverty rates compare across urban and rural areas in the USA?



Scatter plot

There is a clear negative relationship between the percentage of people living in poverty and median house prices. As the percentage of poverty increases, median house prices tend to decrease.

The data points for urban regions are concentrated in the lower poverty percentages (closer to the y-axis), suggesting that urban areas typically have lower poverty rates and higher house prices. Rural areas appear more evenly spread across higher poverty percentages and lower house prices.

Urban regions show a wider spread in house prices, ranging from lower values to as high as \$1.2 million, indicating significant variability in house prices. In rural areas, house prices are generally lower, and variability is more limited.

In areas with over 30% poverty, median house prices are consistently low, regardless of whether they are rural or urban.

1. Economic and Housing Market Disparities:

- The higher median housing prices in urban areas reflect stronger housing markets and economic conditions. Urban areas often have better job opportunities, higher income levels, and greater infrastructure development, driving up demand and prices.

- Conversely, rural areas with lower median housing prices might struggle with economic stagnation, lower demand for housing, and limited access to amenities, which keeps prices down.

2. Poverty and Housing Affordability:

- The lower poverty rate in urban areas suggests better access to economic resources and opportunities that can lift individuals out of poverty. However, the higher housing prices could pose affordability challenges for lower-income residents in urban settings.
- In rural areas, the higher poverty rate indicates more significant economic hardships. The lower housing prices might offer some relief in terms of affordability, but the overall economic struggles likely outweigh this benefit.

3. Urban-Rural Divide:

- These charts highlight the stark differences between urban and rural areas in terms of both economic prosperity and housing markets.

Conclusion/Key Findings

Relationship Between Median Household Income and Median House Prices

- **Strong Positive Correlation:** Median household income and median house prices show a clear positive relationship. Areas with higher household incomes tend to support higher house prices, aligning with the notion that wealthier areas can sustain higher market values.

Employment Rate and Its Impact on House Prices

- **Weak Positive Correlation:** Employment rates exhibit a weak positive correlation ($R = 0.26$) with house prices. While higher employment rates generally indicate better economic conditions, they do not strongly predict house price increases.
- **Concentration in Moderate Ranges:** House prices tend to cluster in regions with employment rates between 85–95%. Beyond these thresholds, house prices show more variability, influenced by other factors like housing supply, external investments, or local market conditions.
- **High House Prices in Low-Employment Regions:** Outliers exist where high house prices persist despite low employment rates. These regions are often influenced by:
 - **Tourism hotspots** (e.g., vacation homes for affluent buyers).
 - **Investment-driven markets** (e.g., real estate speculation).
 - **Luxury and retirement destinations** (e.g., areas catering to high-income individuals or retirees).
 - **Urban housing pressure** (e.g., spillover demand from nearby high-demand urban areas).

Education Rate and Its Impact on House Prices

- We conclude that the relationship between both variables is medium in strength, and that at least a third of the variability in prices can be captured by educational attainment

- The bar chart comparing counties in the top 25% for house prices and education levels showed that roughly 20% of the counties are in both categories, indicating a moderate relationship. This aligns with the R-squared value from the scatter plot, suggesting that educational attainment plays a role in house prices but is not the sole factor.
- The T-test highlighted a significant difference in median house prices between the top and bottom 25% educational attainment counties, reinforcing the impact of education on housing prices.

Poverty Rate and Its Impact on House Prices

- It can be concluded that Higher poverty rates are associated with lower median house prices, indicating an inverse relationship between these variables.
- Urban areas exhibit higher median house prices and lower poverty rates, reflecting stronger economic conditions. In contrast, rural areas face higher poverty rates and lower housing prices, highlighting economic challenges.
- Housing prices in low-poverty areas show greater variability and more outliers, indicating a wider range of property values. High-poverty areas have a more restricted range of housing prices with fewer outliers.
- The analysis underscores a significant economic divide between urban and rural areas, emphasizing the need for targeted economic development and policy interventions to address these disparities.