

# Census Data Analysis

Mikea Mullins

## Introduction

This project aims to examine the relationship between household income and housing value across Connecticut census tracts. This will provide insight into drivers of local property tax revenue, which is essential for funding public services. Here municipal housing value serves as an indicator of a municipality's tax base, and median household income reflects local economic capacity.

## Data

The analysis uses data from the US Census, specifically the median housing value and median household income at the census-tract level in Connecticut.

## Method

A linear regression model was used to investigate the relationship between housing value and household income across census tracts:

$$\text{Median Housing Value} = \beta_0 + \beta_1(\text{Median Household Income}) + \epsilon$$

## Key Findings

- Median household income has a strong positive association with median home value. A \$10,000 increase in median household income is associated with an estimated \$48,070 increase in median housing value ( $p < 0.001$ ). The model explains approximately 64% of variation in housing value across census tracts.
- The model performs best for low- to middle-income census tracts and less well in very high-income tracts.
- High-leverage outliers suggest there are other structural factors influencing housing values. Examples could include population density, housing availability, and proximity to employment centers.

## Limitations

The model does not capture changes over time, and additional unaccounted-for variables may influence housing value (e.g. housing availability, school quality, employment availability). Future models could include time-series data and additional explanatory variables.

## Conclusion

Since property tax revenue is directly linked to housing values, this analysis has several implications for municipal budgeting and policy direction. The high-leverage points, for example, where higher-income census tracts have disproportionately higher housing values, could result in uneven regional fiscal capacity. Moreover overreliance on the tax revenue generated from these high-value tracts could expose municipalities to financial volatility if economic conditions change. In the long term, monitoring growth patterns in income and housing value is useful for predicting future revenue capacity.

## Appendix

### Full regression output:

```
##
## Call:
## lm(formula = house.value ~ hh.income, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -378663  -92742  -17749   54430   957048
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.014e+05  1.219e+04  -8.319  3.7e-16 ***
## hh.income    4.807e+00  1.250e-01  38.471  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 156500 on 815 degrees of freedom
## Multiple R-squared:  0.6449, Adjusted R-squared:  0.6444
## F-statistic: 1480 on 1 and 815 DF, p-value: < 2.2e-16
```

## Graph of linear model

