Census Housing Data

Lorenzo, Elisabeth, An, and Shannon

Question, Interest, & Data Collection

How have housing trends, such as homeownership rates and rental affordability, changed over time across different demographics?

- **Data source**: United States Census Bureau API
- **Exploration**: Over 18,000 variables within the ‘ACS dataset’ based on 5-year survey data
- **Collection**: Identified 22 variables across 4 categories for 2019 - 2022:
  - Household occupancy type, income, and costs.
  - Age
  - Race
  - Poverty status
- **Data Cleaning**: Ensured accuracy by converting data types where necessary and cleaning data to remove missing or incomplete records.
- **Data Loading** : Collected all data from the API and converted to CSV for consistency and collaboration.

```python
# List of Census variables to fetch
variables = [
    "B25003_002E",  # Owner-occupied units
    "B25003_001E",  # Total occupied units
    "B25064_001E",  # Median gross rent
    "B25001_001E",  # Total housing units
    "B19013_001E",  # Median household income
    "B25088_002E",  # Median monthly owner costs
    "B01003_001E",  # Total population

    # Additional age ranges (B25007)
    "B25007_002E",  # Householder under 25 years
    "B25007_003E",  # Householder 25-34 years
    "B25007_004E",  # Householder 35-44 years
    "B25007_005E",  # Householder 45-54 years
    "B25007_006E",  # Householder 55-64 years
    "B25007_007E",  # Householder 65-74 years
    "B25007_008E",  # Householder 75 years and over

    # Additional race categories (B25006)
    "B25006_002E",  # White alone
    "B25006_003E",  # Black or African American alone
    "B25006_004E",  # American Indian and Alaska Native alone
    "B25006_005E"
    "B25006_006E"
    "B25006_007E"
    "B25006_008E"

    # Additional
    "B17021_002E"
]
```

```python
# List of years to collect data for
years = [2019, 2020, 2021, 2022]
```

```python
# Initialize an empty DataFrame to hold combined data
census_combined = pd.DataFrame()
```

```python
# Fetch data for each year
for year in years:
    # Create an instance of the Census library for the specific year
    c = Census(api_key, year=year)

    # Fetch data for the current year
    census_data = c.acs5.get(variables, {'for': 'zip code tabulation area:*'})

    # Create a DataFrame and rename columns
    df = pd.DataFrame(census_data, columns=variables + ['zip code tabulation area'])
    df.rename(columns=columns_mapping, inplace=True)

    # Ensure the ZIP Code is treated as a string
    df['Zip_Code'] = df['Zip_Code'].astype(str)

    # Add the "Year" column
    df['Year'] = year

    # Calculate homeownership rate
    df['Homeownership_Rate'] = (pd.to_numeric(df['Owner_Occupied_Units'], errors='coerce') /
                                pd.to_numeric(df['Total_Occupied_Units'], errors='coerce')) * 100

     # Drop rows with NaN values
    df_clean = df.dropna()

    # Append the DataFrame to the combined DataFrame
    census_combined = pd.concat([census_combined, df], ignore_index=True)
```

Owner_Occupied_Units	Total_Occupied_Units	Median_Gross_Rent	Total_Housing_Units	Median_Household_Income	Median_Monthly_Owner_Costs
3283.0	5509.0	383.0	7250.0	14361.0	771.0
9528.0	12740.0	400.0	17538.0	16807.0	877.0
11118.0	19228.0	433.0	24322.0	16049.0	832.0
1470.0	1946.0	275.0	2783.0	12119.0	526.0
6758.0	8795.0	427.0	12455.0	19898.0	751.0

Homeownership Trend Exploration

Initial Data Exploration

- **Statistical Summary**: Preliminary analysis to understand the distribution, count, and general trends within the data.
- **Yearly Distribution**: Examined homeownership rates by year to identify any trends or significant changes over time.

Variable Selection for Analysis

- **Identified Key Variable:** Focused on a variable that we hypothesized would directly impact homeownership rates: Median Household Income.
- **Relationship Exploration:** Analyzed how Median Household Income affects the homeownership rate.

Analytical Techniques Applied

- **Comparative Analysis:** Used a histogram to show the distribution of homeownership rates across different zip codes and years.
- **Linear Regression:** Employed linear regression to understand and quantify the relationship between homeownership rates and median household income.

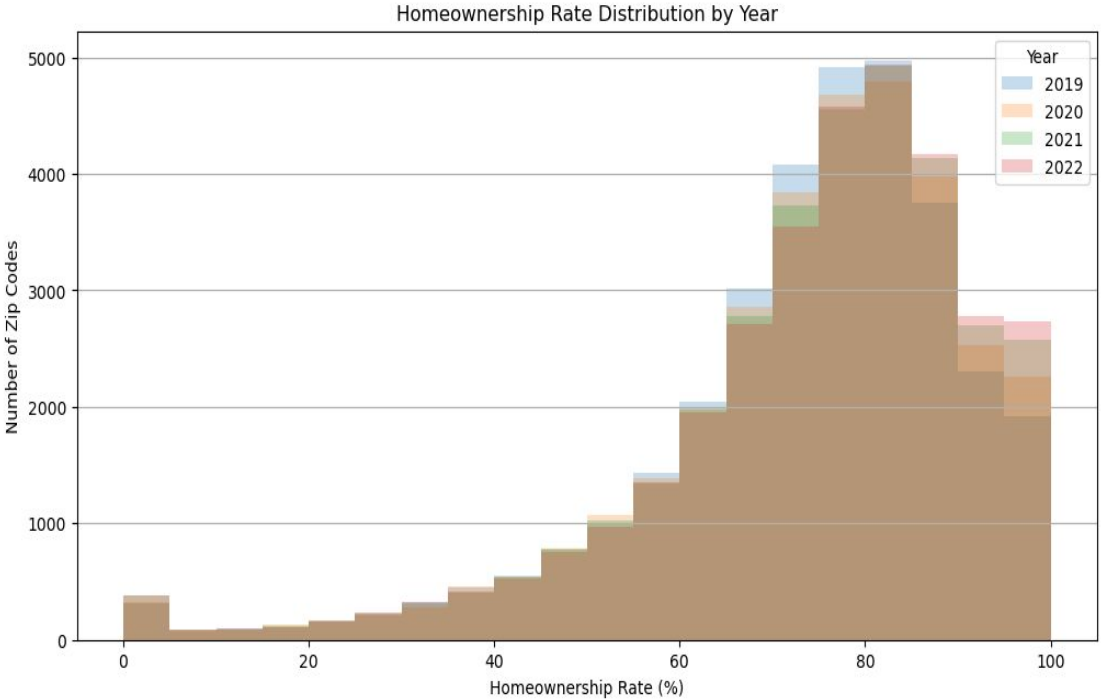
1. How has home ownership rates changed over time?
2. How has household income impacted home ownership rates over time?

Homeownership Rate Distribution by Year

Description
- The graph is a histogram that displays the distribution of homeownership rates across various ZIP codes, segmented by year from 2019 to 2022.

Analysis
- A significant number of ZIP codes have maintained homeownership rates between 60% - 100%
- 2019 showed a slightly higher number of ZIP codes with higher homeownership rates.
- 2022 has shown an increase in ZIP codes where homeownership rates are between 80% - 100%.

Conclusion
- Despite significant economic events between 2019 and 2022, the distribution remained similar, pointing to resilient homeownership trends.



Impact of Household Income on Homeownership Rates

Description

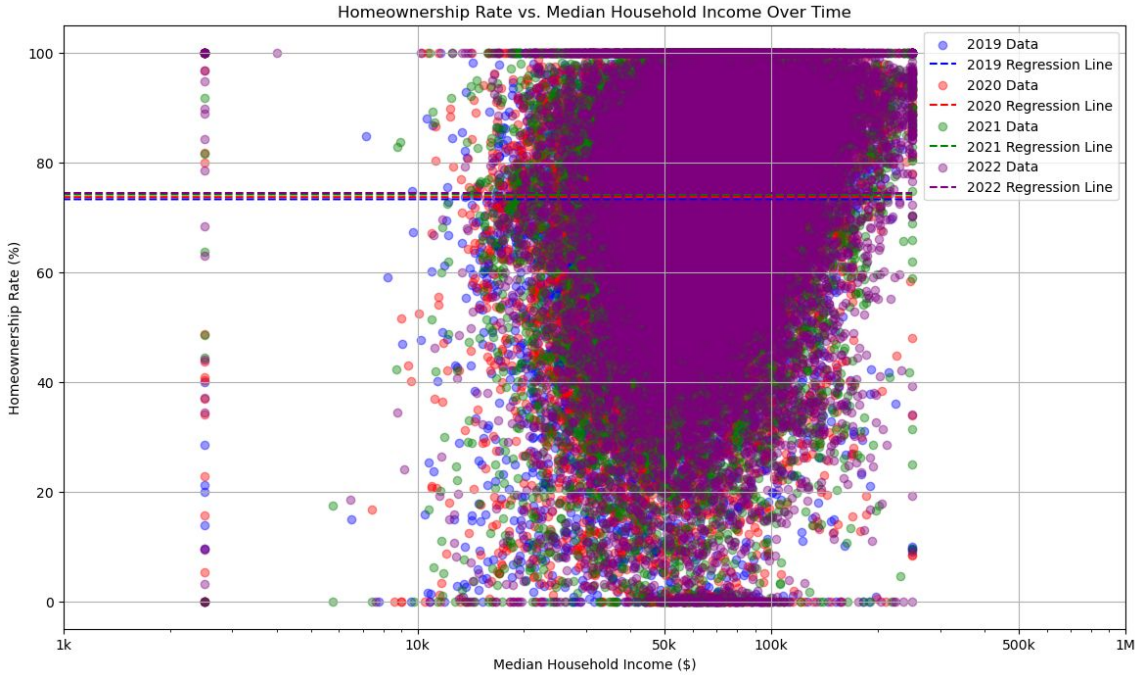
- A scatter plot with regression lines for each year, illustrating the relationship between median household income and homeownership rates between 2019 -2022.

Analysis

- Consistent negative slope across all years suggests higher incomes don't necessarily predict higher homeownership rates. Implying higher income does not lead to increased homeownership, possibly due to higher property prices or preferences for renting.
- For each year, the p-values are extremely low (approaching 0) and the regression lines are nearly horizontal, suggesting that other factors might be influencing homeownership rates more strongly than median income alone.

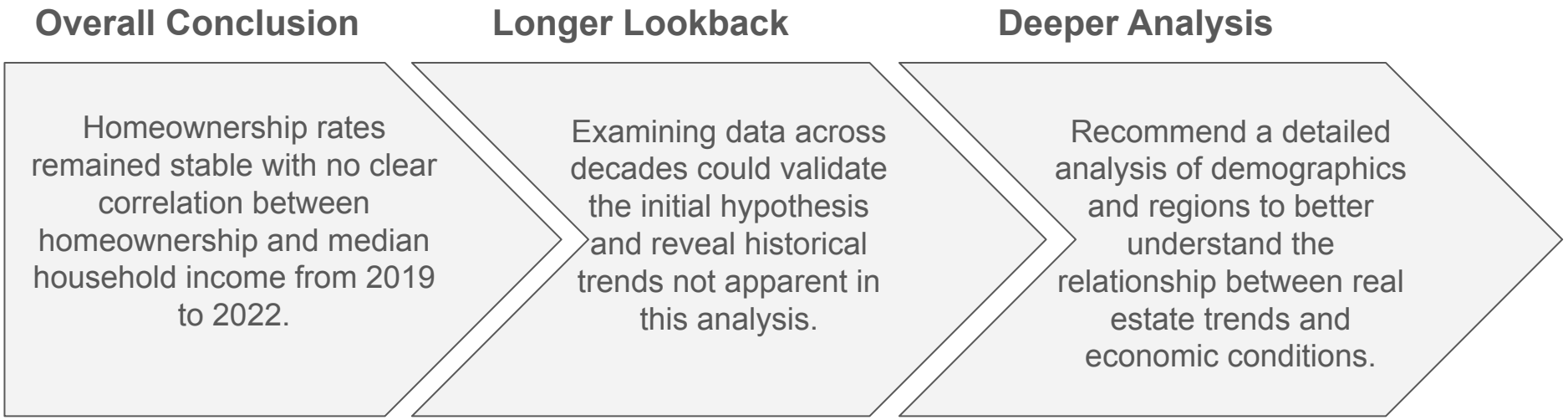
Conclusion

- While hypothesized that median household income is a major factor in homeownership, the analysis indicates a more complex relationship that varies across ZIP codes and time.



Year	Slope	Intercept	R-Squared	P-Value
2019	-2.468	73.271	0	0
2020	-2.808	73.664	0.001	0
2021	-1.77	74.149	0	0.004
2022	-1.219	74.411	0	0.046

Homeownership Conclusion and Takeaways



Regional Analysis Questions

National	Midwest	Three State	Individual
Median gross rent vs population	Do midwest zip codes display the same relationship of zip code population to median gross rent?	Can the same relationship be represented by comparing 3 states from differing region?	Do independent state zip codes show the same relationship between zip code population and median gross rent as was seen in larger analyses?
Poverty percentage vs population	How do the median rents of each state compare to each other?	East Coast: Connecticut Midwest: Minnesota West Coast: California	
Poverty percentage vs median gross rent			

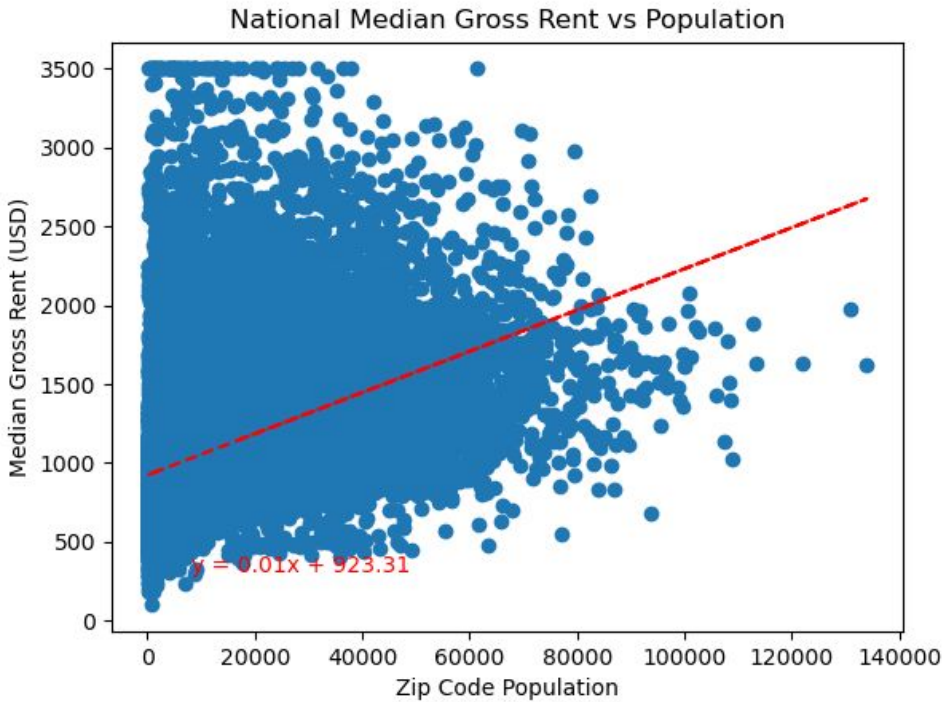
National Analysis

- Median gross rent vs population
- Poverty percentage vs population
- Poverty percentage vs median gross rent

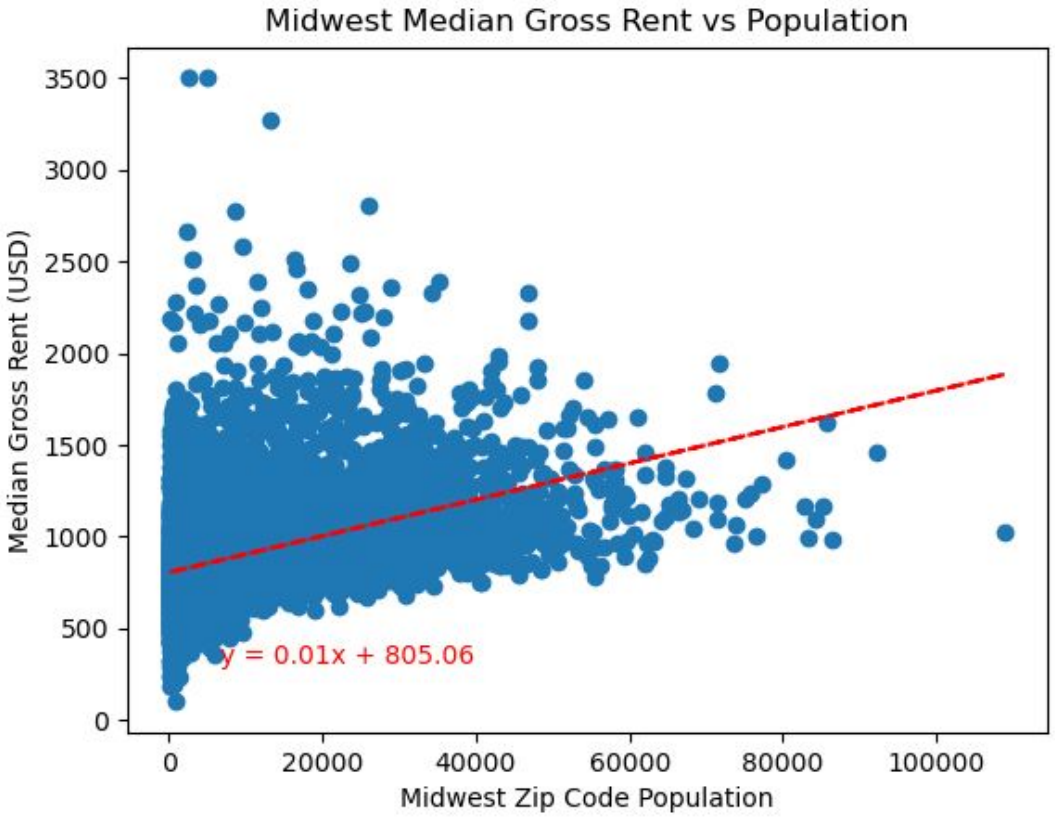
Assessed P and R values to select a relationship for further analysis

P = 0.0

R = .432105



Midwest Analysis



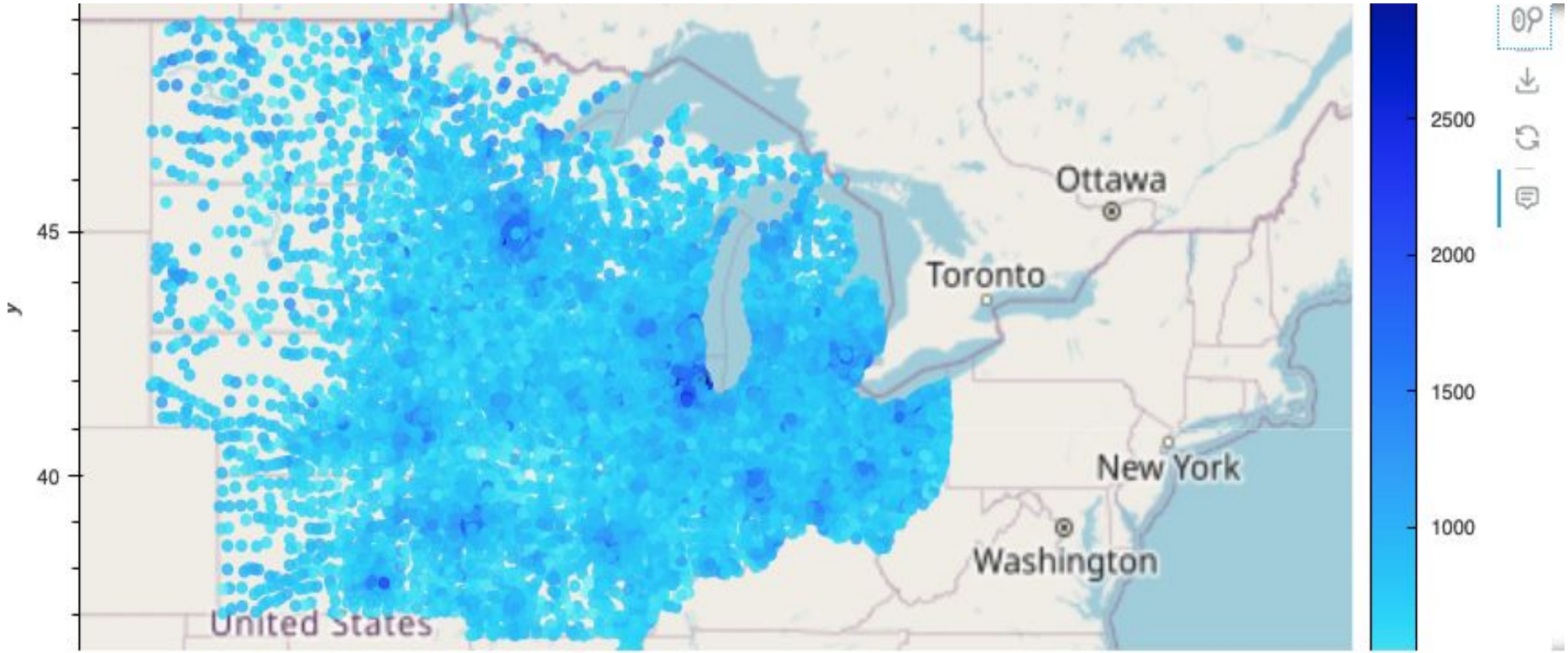
Moderate Positive Correlation

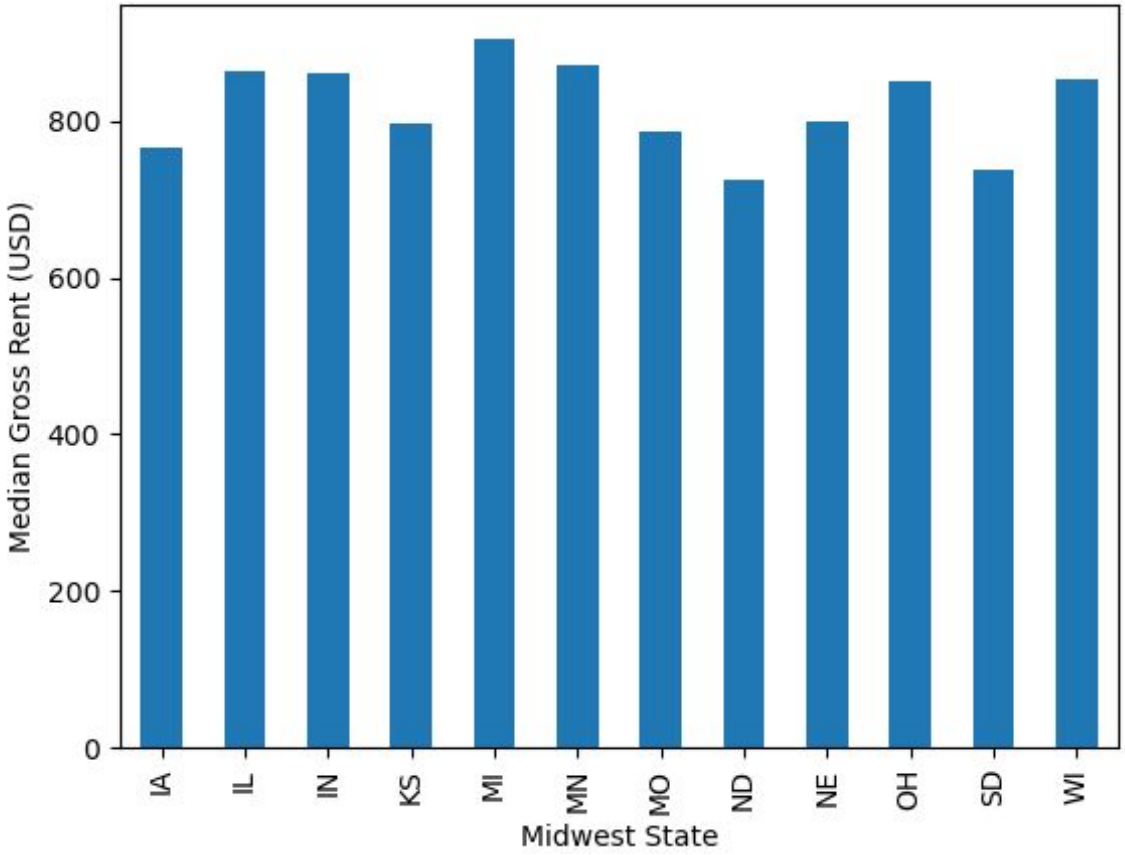
R = .431088

P = 0.0*

Zip Code Ranges

Mapped Midwest MGR



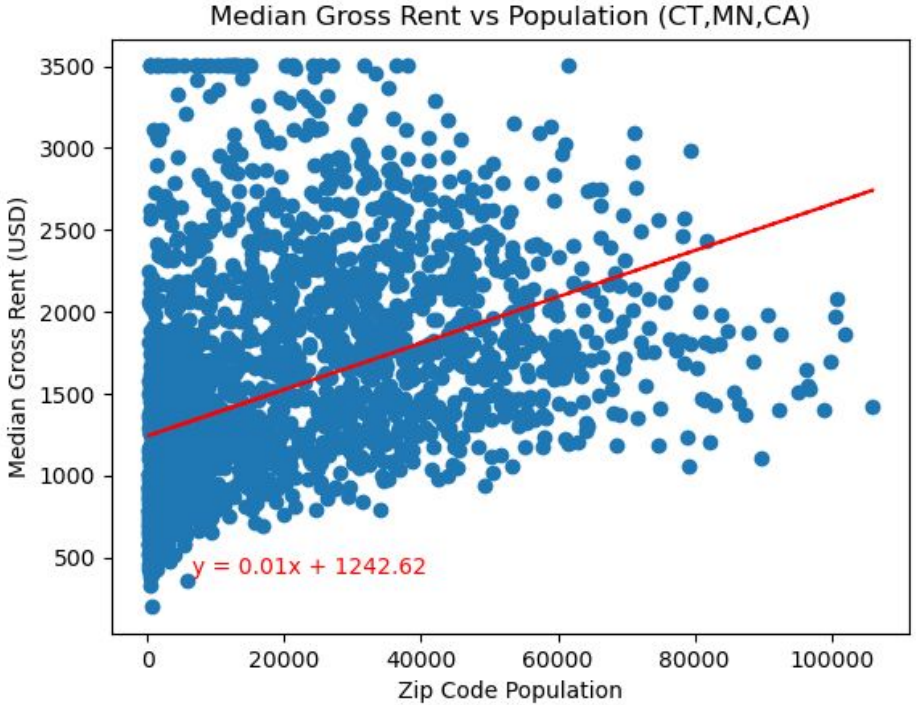


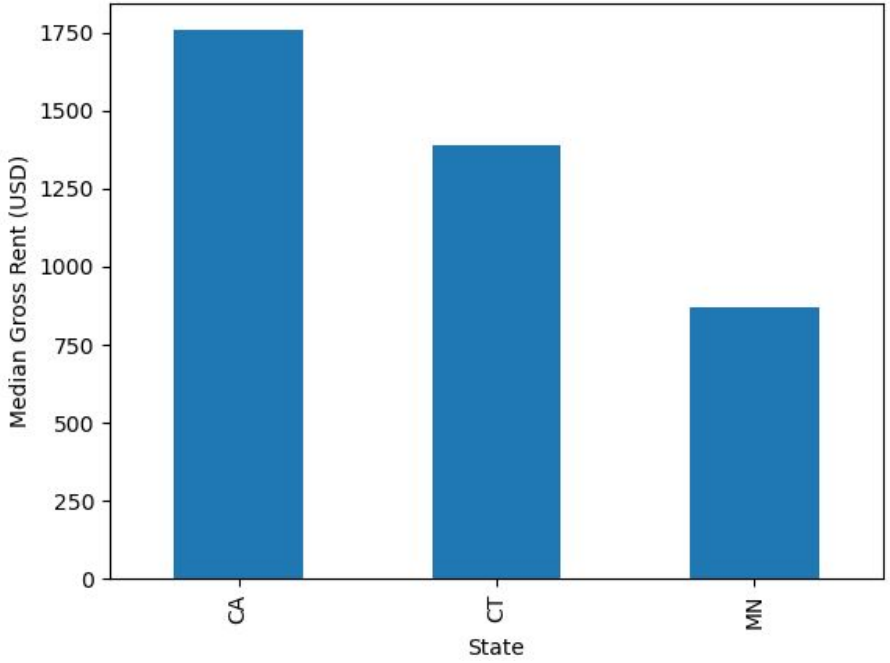
MEDIAN GROSS RENT
By State

STATE MAX:
MI $904

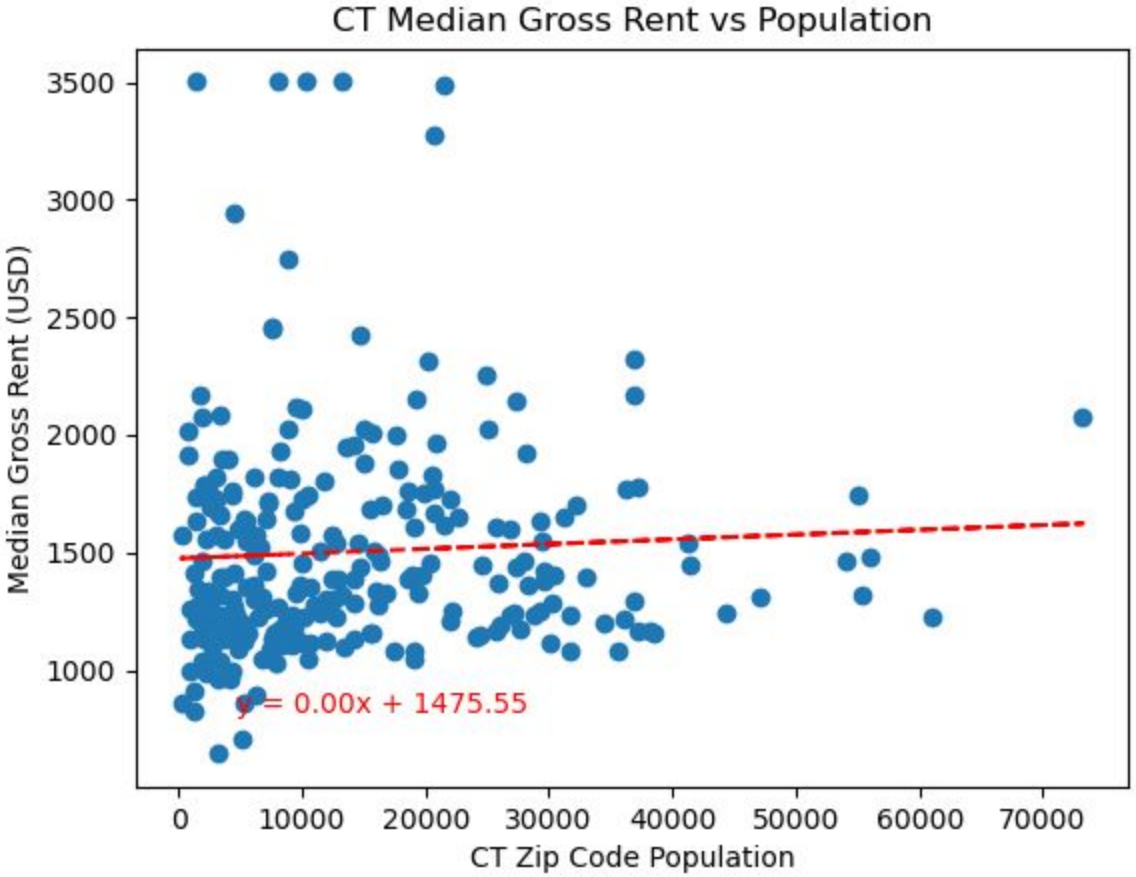
ZIP CODE MAX:
IL (2) $3,501*

Three State Analysis





Individual State Analysis



Connecticut MGR vs Pop

P = .39

Fail to Reject the Null

Hypothesis Testing: Independent T-tests

MN vs CT

Null: There is no difference in median gross rent between Minnesota and Connecticut zip codes

REJECT THE NULL

P = 5.713259**e^-277**

MN vs CA

Null: There is no difference in median gross rent between Minnesota and California zip codes

REJECT THE NULL

P = 9.54239**e^-48**

CA vs CT

Null: There is no difference in median gross rent between California and Connecticut zip codes

REJECT THE NULL

P = 4.1308456**e^-17**

Affordability Trends and Demographics

Rent/Income

How has the average rent//income ratio changed from 2019-2022

Owner Cost/Income

How has the average owner cost/income ratio changed from 2019-2022

By Age

Do different age groups experience similar trends in rent/income ratio

Average Rent/Income Ratio by Year

Description:

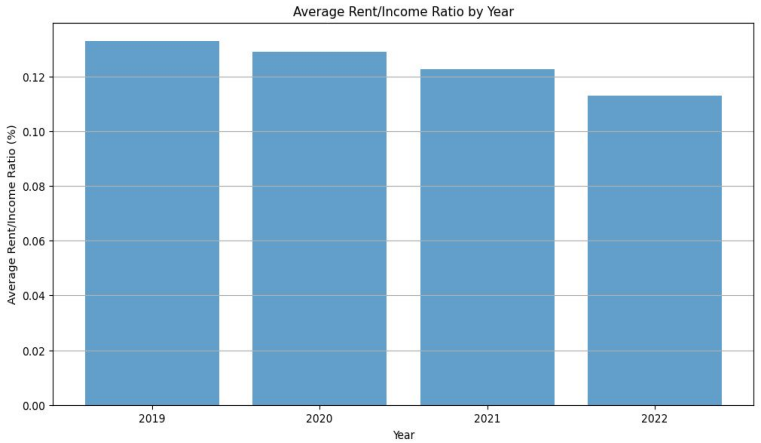
- Bar graph showing average rent/income ratio by year (2019-2022)

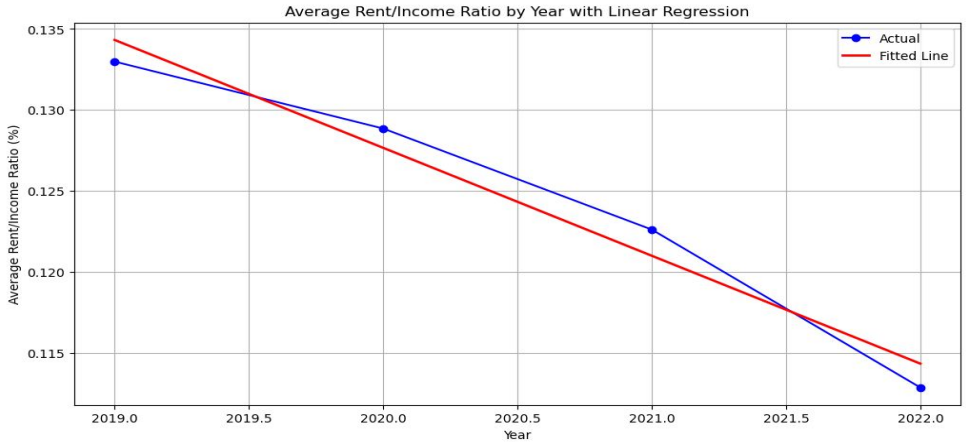
Analysis:

- Rent income/ratio is relatively stable from 2019 to 2021 hovering around 12%
- Noticeable decline in 2022, where the ratio drops to just about 10%

Conclusion

- On average, households spend about 12% of their income on rent and dropped to about 10% in 2022
- Economic policies or increase in median household income could be potential reasons for the decline in 2022





Average Owner Cost/Income Ratio by Year

Description:

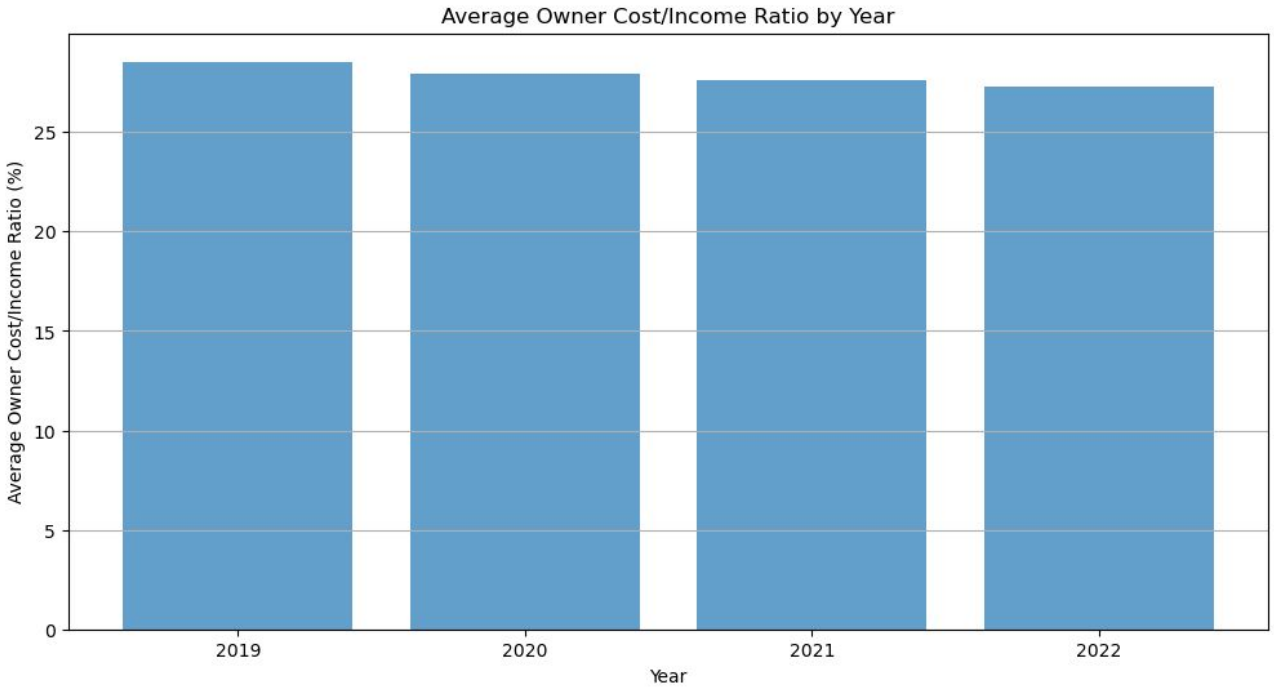
- Bar graph showing owner cost/income ratio by year (2019-2022)

Analysis

- Owner cost/income ratio is consistent from 2019 to 2022, on average households spend about 25% of their income on ownership cost
- Stability suggest that ownership cost, relative to income, have remained consistent over these years

Conclusions

- Stability in owner cost/income ratio could be due consistent property taxes rates, fixed mortgage rates, or stable maintenance cost
- Could also indicate median household income has increased at a rate that matches with increase in ownership cost



Average Rent/Income Ratio by Age Group

Description

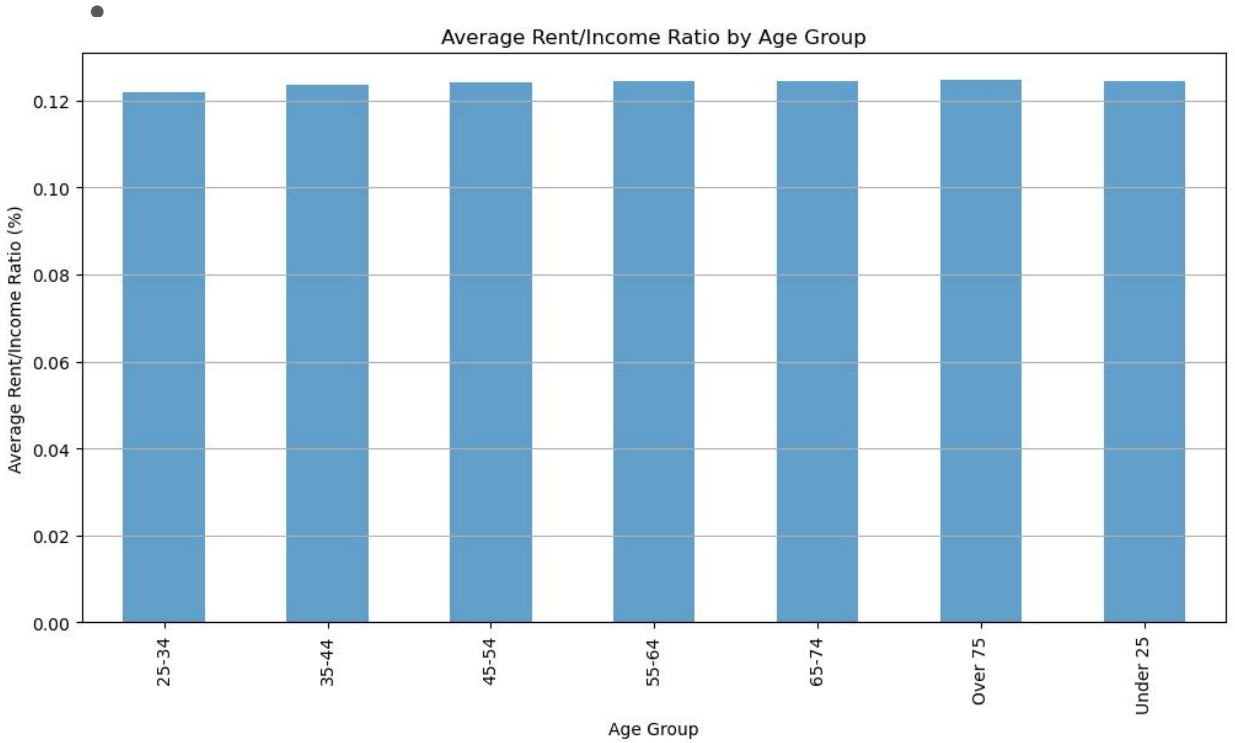
- Bar graph showing average rent/income ratio by age group

Analysis

- The average rent/income ratio is consistent across all age groups with each age group having a ratio around 12%

Conclusions

- Households, regardless of age are spending a similar proportion on rent



Average Owner Cost/Income Ratio by Age Group

Description:

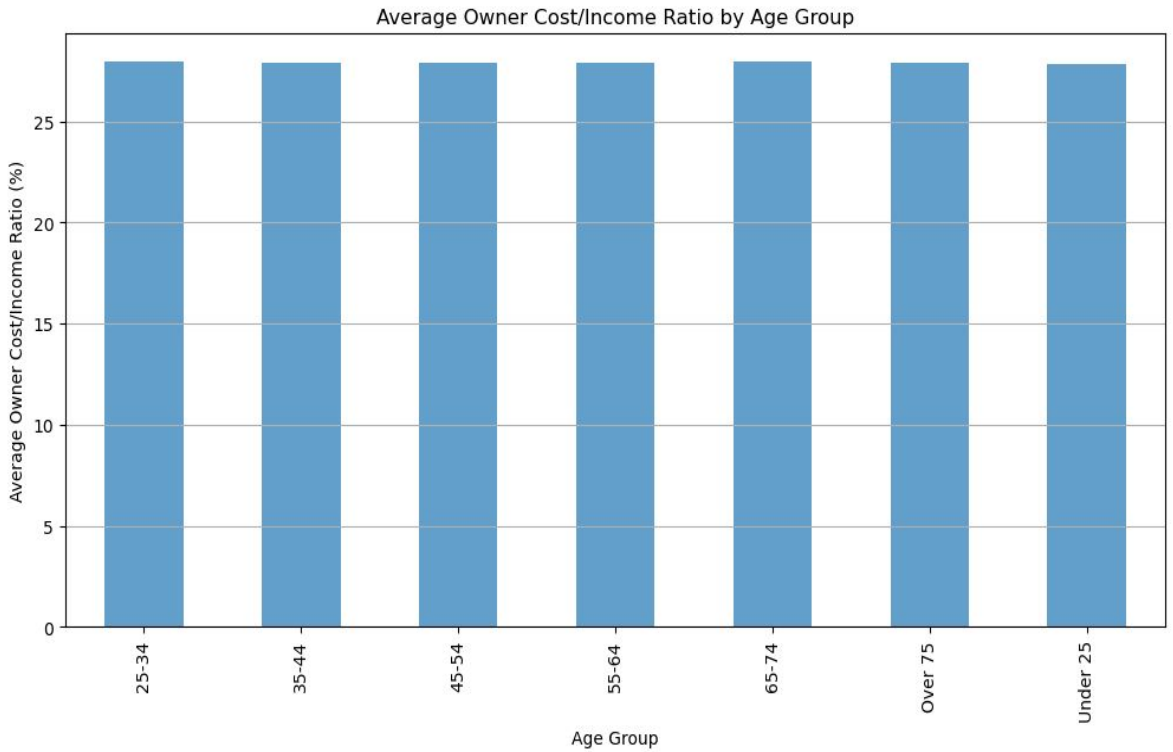
- Bar graph showing average owner cost/income ratio by age group

Analysis:

- No significant variation among age groups
- Each age group had a ratio around 25 %

Conclusion:

- No specific age group is disproportionately affected by ownership cost



Conclusion

Average Rent/Income Ratio by Year and Average Owner Cost/Income Ratio by Year	Average Rent/Income Ratio by Age Group and Average Owner Cost/Income Ratio by Age Group	Deeper Analysis
Average rent/income ratio through 2019-2022 is relatively stable with a slight change in 2022 while average owner cost/income is consistent and stable through from 2019-2022	Average rent/income ratio is consistent throughout all age groups. There is also no significant variation among age groups for average owner cost/income ratio	Deeper analysis of earlier years and different demographics/ factors can help to identify more trends

Homeownership and Affordability: Trends over Time

Data Collection: Selected data from Census as a group - acs5. Collected cost of living information from US Bureau of Labor Statistics and American Institute for Economic Research.

Variable Selection: Working with the group, selected appropriate variables for larger project. Added specifics for demographic analysis.

Analytics: Visualization with bar and line graphs; calculated R-square and p-score; and used excel for basic mathematics.

Homeownership and Affordability

By income	How has median household income changed over time?
By gross rent	How has median gross rent changed over time?
By owner costs	How has median monthly owner cost over time?
By race, by ownership	How has the relationship between race and ownership changed over time?

What is the relationship between median household income, median gross rent, median monthly owner costs, and race demographics?

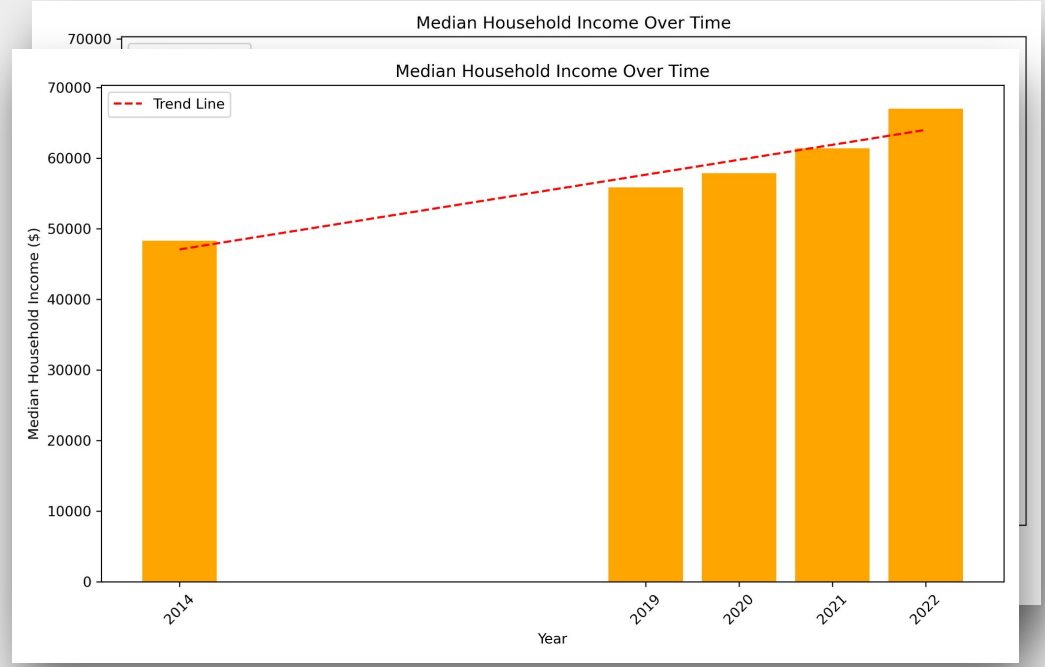
Homeownership and Affordability: Median Household Income/Time

Bar graph showing median household income 2019-2022 and 2014-2024.

2019-2022 trend line with R-square = 0.9562

2014-2022 trend line with R-square = 0.9078

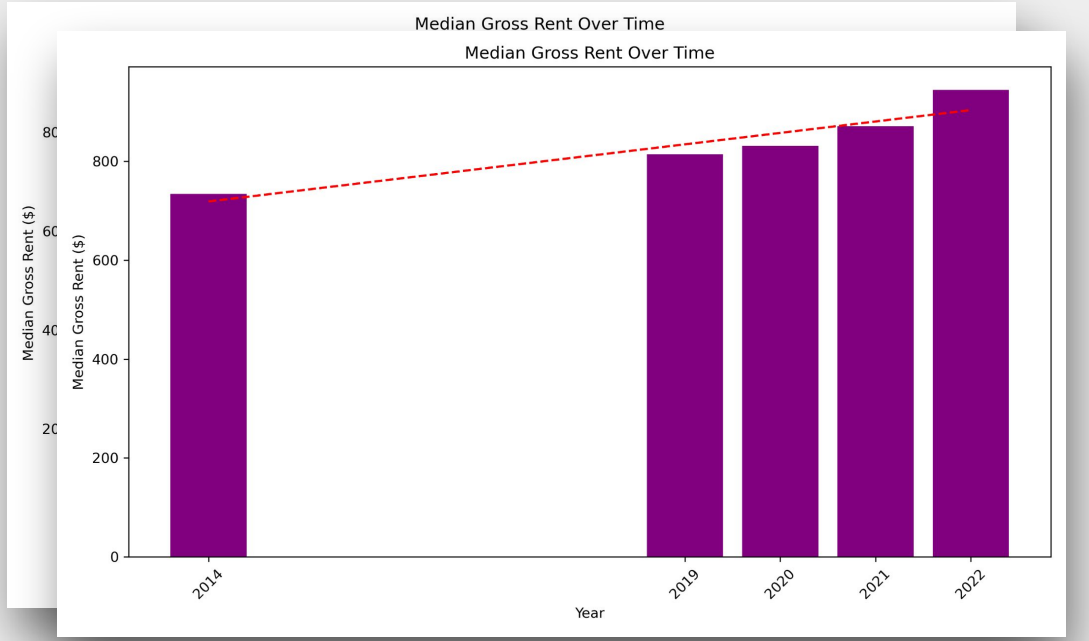
Median household income has increased consistently over time - both r-squared values indicate high and reliable upward trend.



Percent change from 2019-2022 = 19.98%

Percent change from 2014-2022 = 38.69%

Homeownership and Affordability: Median Gross Monthly Rent Cost/Time



Bar graph showing median gross monthly rent 2019-2022 and 2014-2024

2019-2022 trend line with R-square = 0.9214

2014-2022 trend line with R-square = 0.8711

High R-squared values provide statistical confirmation of the visual upward trends observed in both charts

Percent change from 2019-2022 = 15.97%

Percent change from 2014-2022 = 28.61%

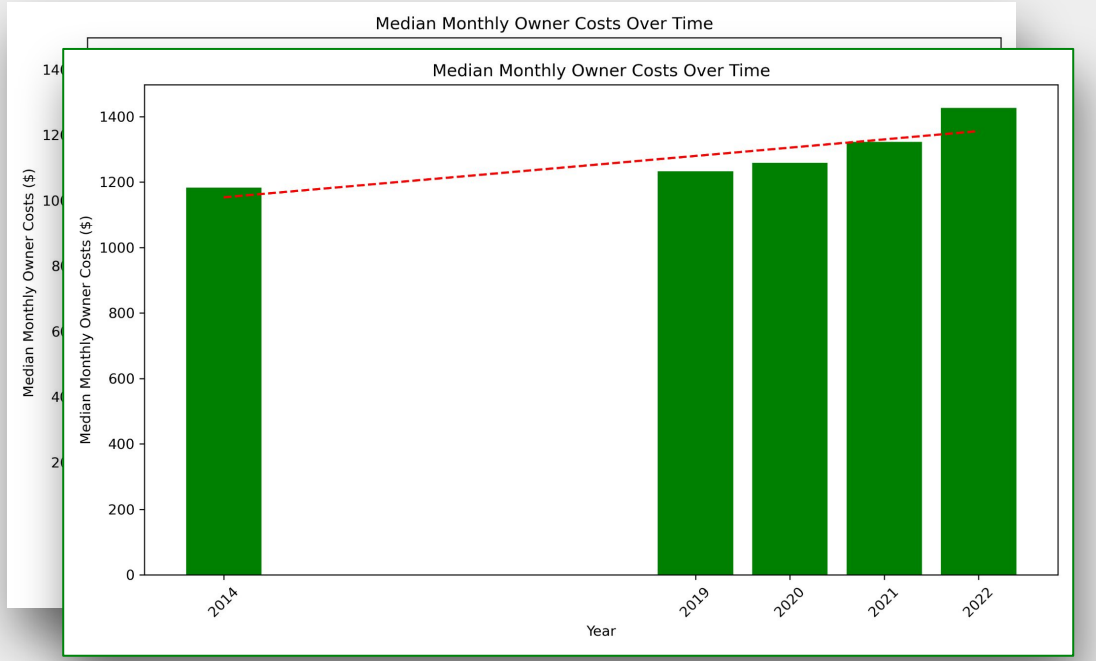
Homeownership Affordability: Monthly Ownership Cost/Time

Bar graph showing median monthly ownership costs 2019-2022 and 2014-2024.

2019-2022 trend line with r-square = 0.9301

2014-2022 trend line with r-square = 0.7095

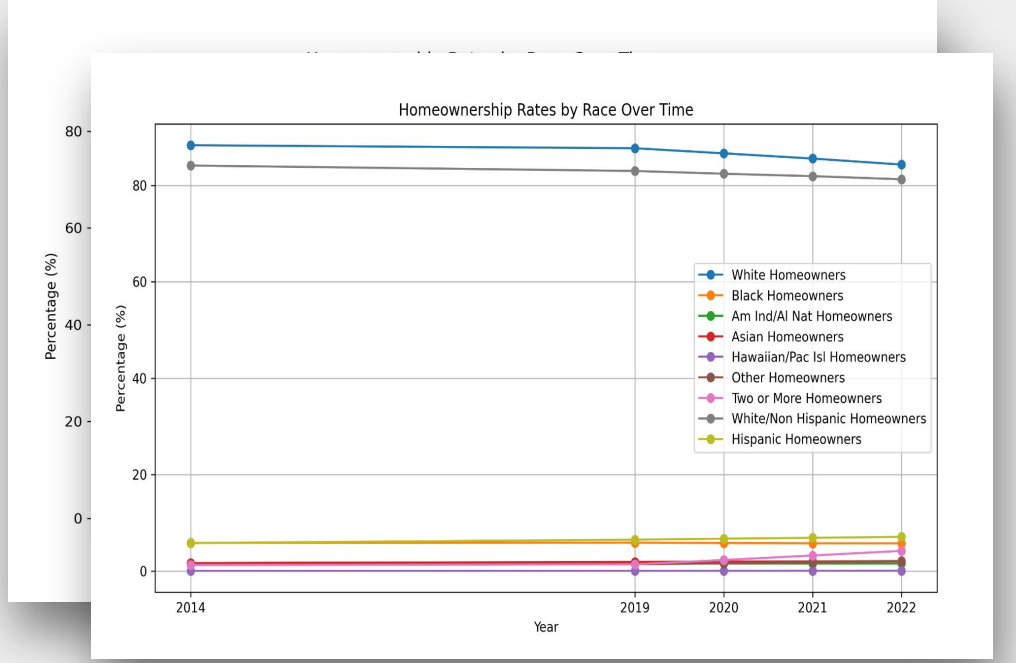
R-square values provide clear indications of short-term stability, longer-term variability.



Percent change from 2019-2022 = 15.61%

Percent change from 2014-2022 = 20.54%

Homeownership and Affordability: Race and Ownership Trends Over Time



Line graph showing race and ownership across 2019-2022 and 2014-2024

Analysis of 9 racial groups to find trends in rate over time. Determined R-squared and p-value.

Able to reject the null for 5 of the 9 groups (Asian, Hawaiian/Pac Islander, “Other”, White/Non-Hispanic,and Hispanic.)

White/Non-Hispanic homeowners have shown a statistically significant decrease in homeownership rates; Asian, Hawaiian/Pacific Islander, Other, and Hispanic homeowners have shown significant increases in homeownership rates.

Homeownership and Affordability: Are we keeping up?

Consistently Rising Costs: Both ownership and rental costs have been rising, creating affordability challenges.

Income Growth: Median household incomes are increasing, which helps mitigate some of the affordability issues but may not be sufficient for all groups.

Racial Disparities: Significant trends in homeownership rates vary by race, highlighting disparities in access to homeownership that need attention.

Cost of living (bls.gov):
From 2019-2022 = 11.69%
From 2014-2022 = 20.69%

Cost of living (aier.org):
From 2019-2022 = 14.47%
From 2014-2022 = 23.62%

Major Findings

- Homeownership rates have remained relatively constant between 2019-2022, suggesting that we need to look back further to identify any trends.
- Despite economic factors and events between the years of 2019-2022, we don't see a correlation between median household income and homeownership. This suggests that there are likely other factors that impact homeownership.
- Both short-term and long-term analyses show a strong upward trend in monthly ownership costs, indicating that owning a home is becoming more expensive over time. This trend is slightly more variable over the long term, suggesting fluctuations in economic conditions or housing policies.
- Similar to ownership costs, median gross monthly rent costs have also been increasing consistently. This trend highlights the growing affordability challenges in the rental market, affecting those who may not be able to afford homeownership.

Next Steps

- Homeownership rates have remained relatively constant between 2019-2022, suggesting that we need to look back further to identify any trends.
- The strong upward trend in median household income suggests that, on average, households are earning more over time. However, the increase in household income must be compared against the rising costs of homeownership and rent to understand overall affordability.
- Zip code population, in many cases, is a factor in median rent for the area. Further exploration can be done into what is bringing people to specific zip codes and what other factors are contributing to rental costs.
- Average rent/ income ratio and average owner cost/income ratio is relatively consistent between 2019-2022. Average rent income ratio and average owner cost/income across different age groups is also relatively consistent, looking back at earlier years and different demographic factors can help identify any trends

Questions?