Unit 35.1 - Capstone Three – Final Project Report
Gayla Rios

Housing Affordability and Debt in America

1. Problem Identification
Investigate the relationship between housing affordability (both in terms of rent and home value) and overall debt at a county level in the United States. We aim to understand whether counties with higher housing costs also experience elevated levels of debt, thereby providing insights into the financial challenges faced by residents in different counties.

With the ever-fluctuating real estate market and the diverse economic challenges faced by different regions, understanding the correlation between housing costs and debt can offer policymakers, financial institutions, and community organizers crucial insights. Housing is a significant expense for most households, and understanding its impact on overall financial health is pivotal.

2. Data Sourcing
1. **Urban Institute Debt in America - County Level Overall Debt**
Contains information on overall debt levels for each county, with data categorized by various factors. Urban Institute Dataset

2. **Zillow Observed Rent Index by County**
Provides insights into the rental market, indicating housing affordability in terms of renting. Zillow Research Data

3. **Zillow Home Value Index by County (Mid-value homes in 35th – 65th percentile)**
Reflects the average home values for homes in 35th-65th percentile, serving as a metric for housing affordability for home buyers. Zillow Research Data

3. Data Cleaning
1. I had to remove all the columns that segmented types of debt in the Urban Institute dataset leaving only the totals for 'Auto/Retail Loan Delinquency', 'Credit Card Delinquency', 'Medical Debt in Collections', and 'Student Loan Default' as the segmented columns for each of those totals had excessive missing values.
2. Removed all characters after "," in 'County' in Urban Institute dataset so that it would be recognized as a geographic field type in Tableau.
3. Merged Urban Institute and both Zillow datasets together using and inner join on 'StateFIPS' and 'CountyFIPS'.
4. Since the Urban Insitute dataset values are the average of the past year's values as of September 2023, I dropped all the columns in both Zillow datasets (home values and rent) except 2023 columns. I then took an average of the 2023 columns and created 'Average Zillow Home Estimate and 'Average Zillow Rent Estimate' columns.
5. Dropped all 'FIPS' as they were no longer needed. Dropped redundant columns leaving only one column for features such as county, state, and region.
6. After all the above, there were only 3 null values in 'Median Debt in Collections' and 7 null values in 'Student Loan Debt in Default' out of a total of 1,058 rows. With so few, I decided to drop those rows – removing a total of 8 counties from the dataframe (turns out only 8 rows were removed eventhough there were 10 missing values)
7. Simplified and shortened all column names.
8. Reordered columns to organize.
9. Rounded float columns to 3 digits.
10. Changed whole number columns to integer types.
11. Saved cleaned dataframe to Data/data_clean.csv

4. Exploratory Data Analysis
I started by first plotting histograms for all numeric columns. This uncovered two significant outliers in the 'Average Zillow Home Estimate' and 'Average Zillow Rent Estimate' columns. I sorted those columns and discovered that Pitkin County, CO and Teton County, WY had excessively high values. Both of those counties include cities where billionaires vacation (Aspen, CO and Jackson Hole, WY).

# Unit 35.1 - Capstone Three – Final Project Report
Gayla Rios

Since the estimates seemed accurate as opposed to typos, I decided to leave them in the dataframe. I'm not sure if that was actually the best decision because in the end, I repeatedly had to filter them out to see relationships in the data. It probably would have been best to remove them.
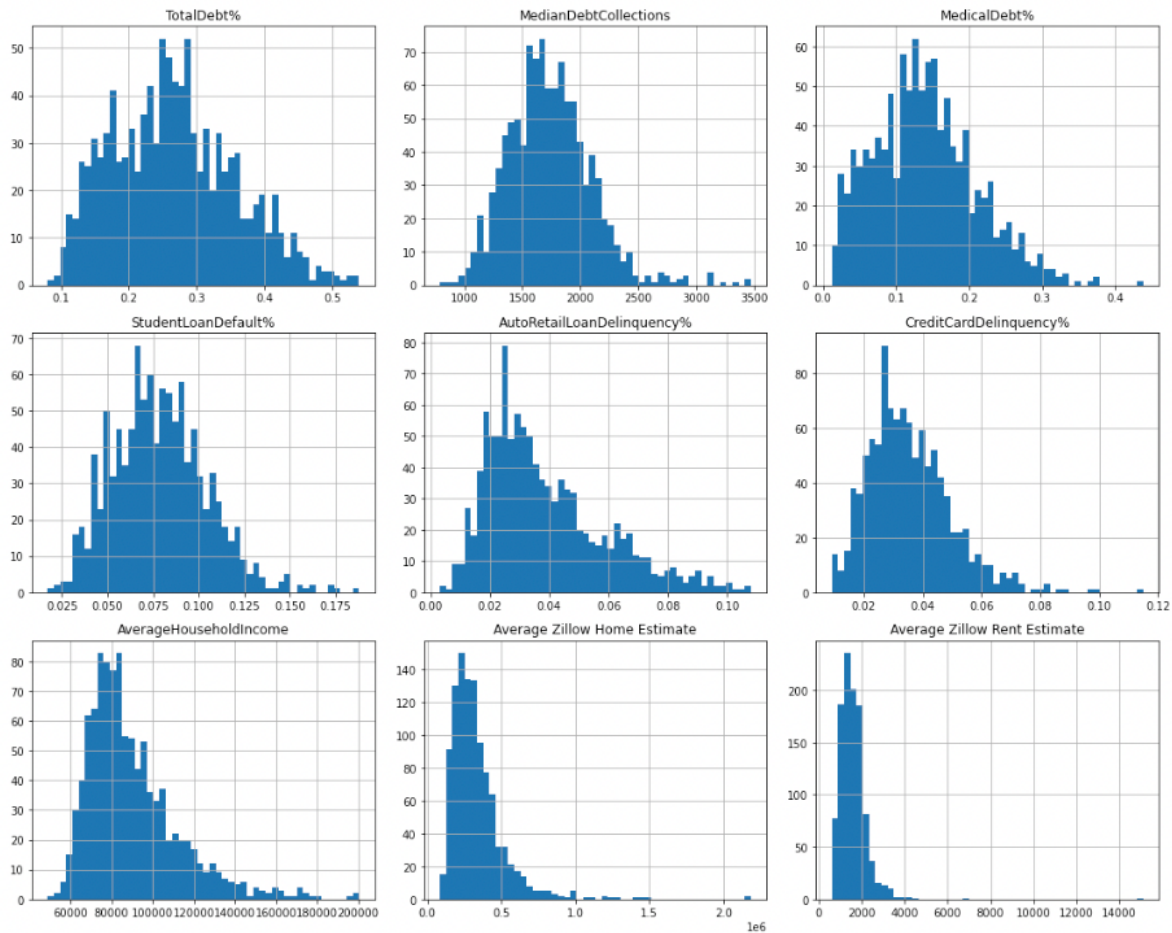
Unit 35.1 - Capstone Three – Final Project Report
Gayla Rios

I also created a heatmap to see how variables correlate. Income and debt variables were inversely correlated. Zillow estimates for home values and rents were also inversely correlated.
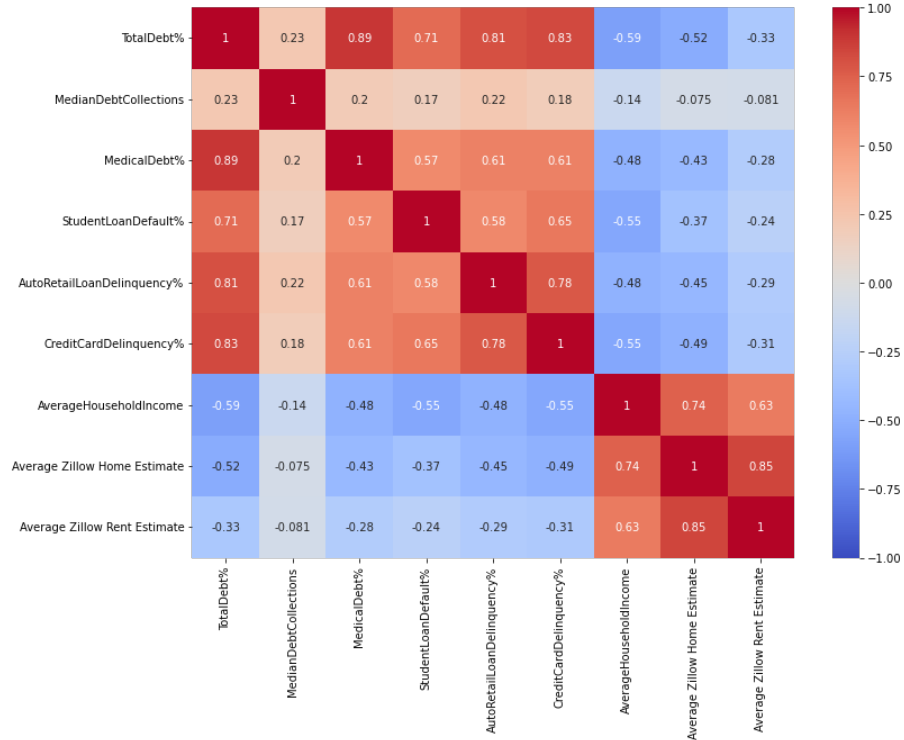
I then created a scatter plot to explore the relationship between 'Average Zillow Home Estimate', 'Average Household Income', and the population 'TotalDebt%'.
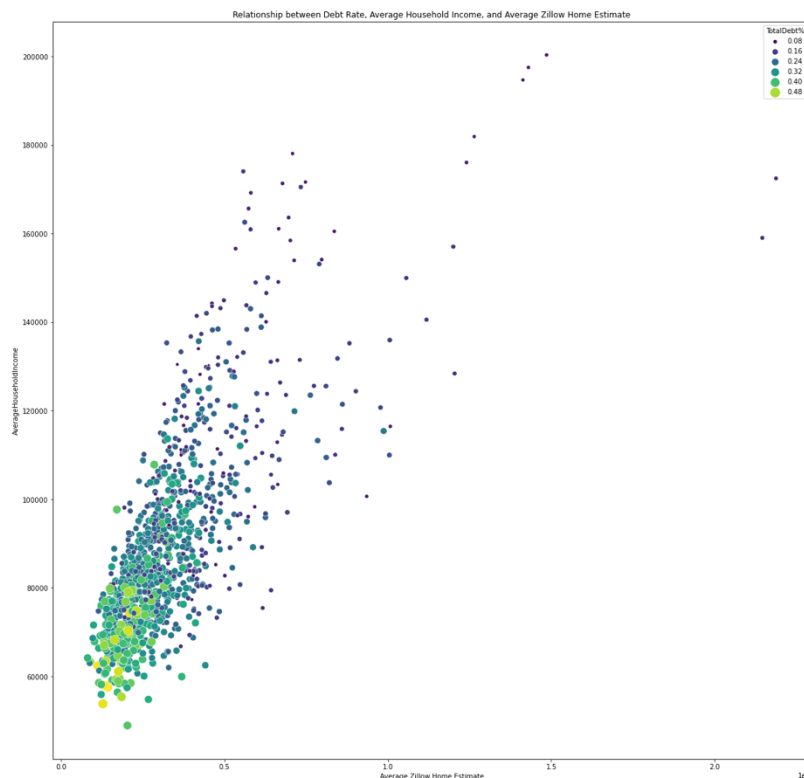
As average household income increased in each county, so did home values. Those communities with the highest rates of debt have the lowest home values and household incomes.

5. Tableau Story Points Housing Affordability and Debt in America
Beginning with a state-by-state analysis of household income versus debt trends. The top graph showcases the average household income juxtaposed with the total percentage of the population in debt for each state. The lower graph delves deeper, segmenting the debt rates into categories: Auto/Retail Loan Delinquency, Credit Card Delinquency, Medical Debt in Collections, and Student Loan Default. Notably, while some states exhibit high household incomes, they also have significant debt rates, suggesting that a higher income does not necessarily equate to lower debt levels. Conversely nearly all states with highest debt levels have lower income levels.

I next assessed the impact of how debt patterns influence housing affordability across U.S. counties.


The left chart below illustrates the correlation between the percentage of the population with debt in collections and their housing affordability, emphasizing the potential burden of accumulating debt on maintaining a roof over one's head. Meanwhile, the right chart underscores the link between the prominence of medical debt within total debt collections and housing costs, shedding light on the potential financial repercussions medical emergencies can have on housing stability and affordability.

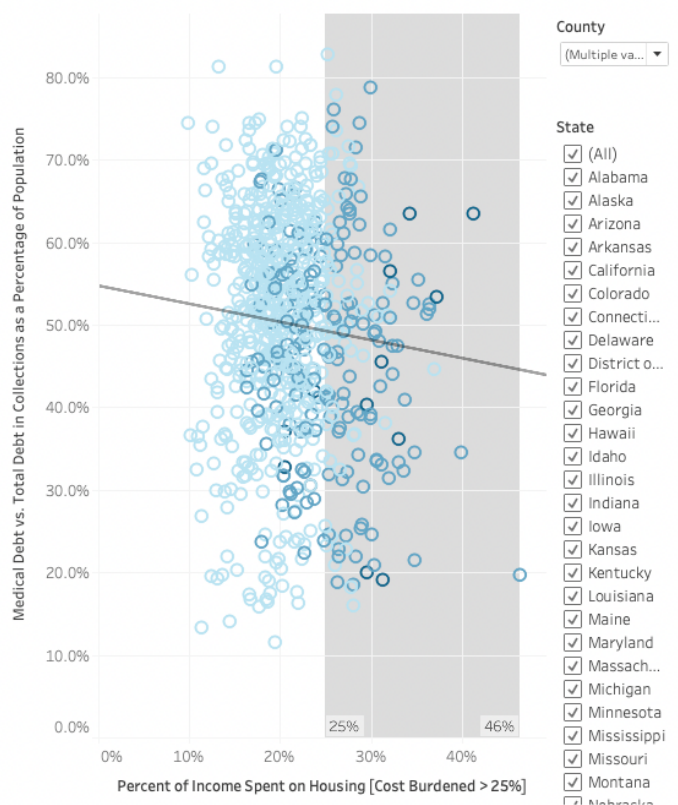# Unit 35.1 - Capstone Three – Final Project Report
Gayla Rios

These scatterplots delve into the intricate relationship between debt metrics and the ability for
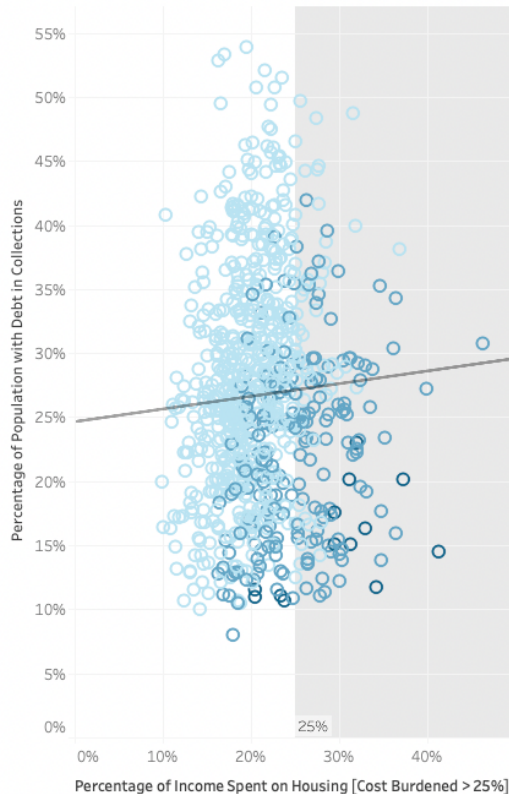
residents to manage housing costs. Since I only have rent estimates, I chose to highlight a threshold of 25% to account for standard 30% Rule adjusted by additional 5% of income for housing related costs. Essentially people spending more than 30% of their income on housing related expenses are cost burdened. I broke this down further by the ratio of 'medical debt rate'/'total debt rate' as medical debt is the highest proportion of total debt in nearly all states. We see a slightly inversed relationship indicating in areas (or among groups) where a higher proportion of debt is due to medical reasons, a smaller proportion of income is spent on housing this might be an indication of financial pressure due to higher medical debt. This would be an excellent area for further study.

Unit 35.1 - Capstone Three – Final Project Report
Gayla Rios

I next looked at a comparison of housing costs and debt trends by county and state. The top graph below presents a county-level comparison of the median debt in collections against Zillow's estimated rent costs, revealing patterns and disparities across various counties. Meanwhile, the bottom graph illustrates a state-level overview, juxtaposing the average household income with Zillow's home estimates. The visualization underscores the intricate relationships between income, housing costs, and debt, highlighting potential areas of concern or stability across the US.
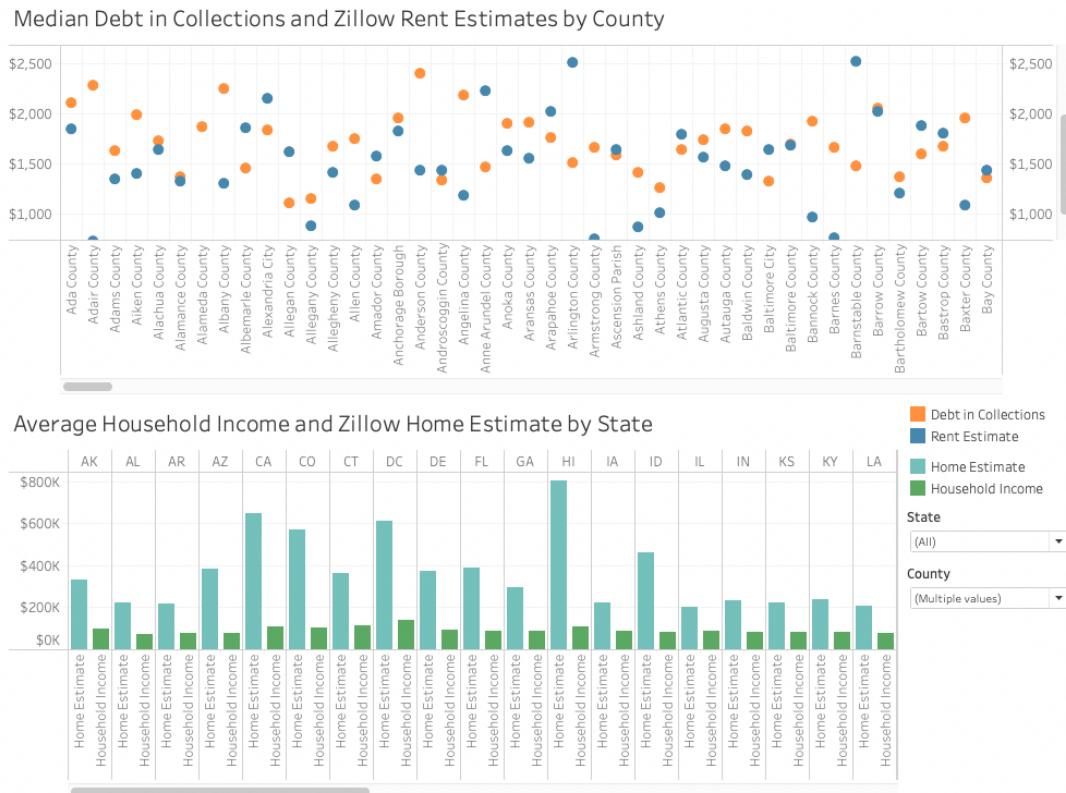

Let's now look at unraveling the ties between income, home values, and debt. The left chart below depicts the positive correlation between household incomes and Zillow home estimates, with each point colored based on the county's population debt in collections rate. A discernible upward trend suggests higher home values in counties with higher incomes. On the right, the inverse relationship between the percentage of a county's population with debt in collections and its average household income is portrayed. The gradient of color intensity represents the proportion of the population with debt, illuminating potential socioeconomic stress points across different counties.

Relationship between Household Income and Zillow Home Estimate by County (Colored by Population Debt in Collections Rate)

Relationship between Debt in Collections Rate and Household Income by County
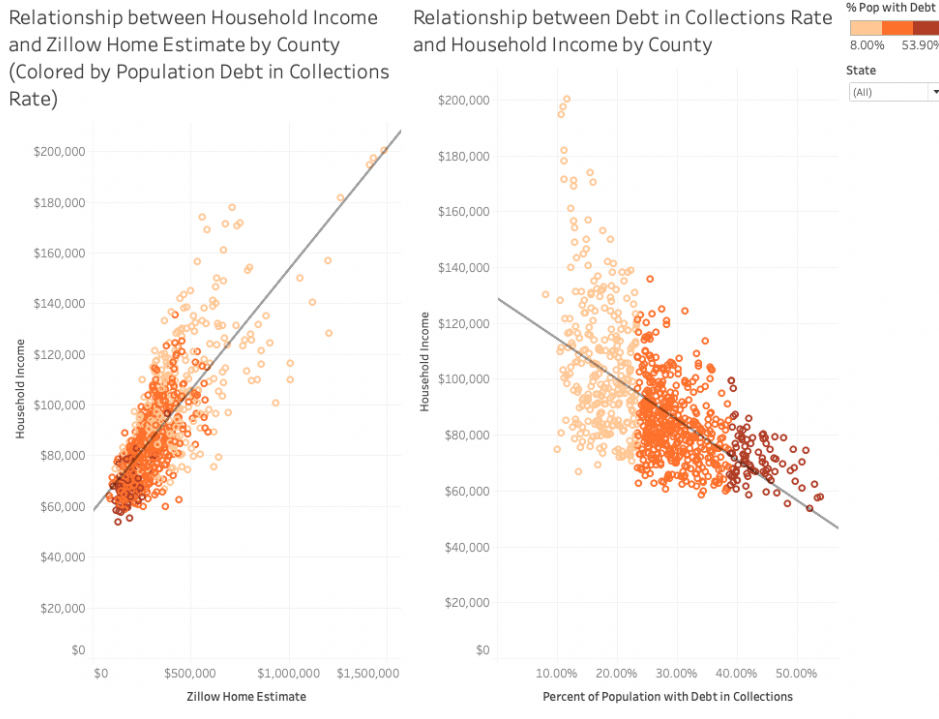
Lastly we'll examine financial strains across U.S. counties. The left map illustrates counties based on their population's debt in collections rate, with darker shades indicating higher rates, signifying areas with increased financial burdens. On the right, the focus shifts to the annual rent-to-income ratio by county. Lighter shades represent counties where residents spend a smaller portion of their income on rent, potentially indicating greater housing affordability, while darker shades highlight areas where rent takes up a significant chunk of household earnings. Together, these maps shed light on regional variations in financial challenges faced by residents.

Relationship between Debt in Collections Rate and Household Income by County

Annual Rent to Income Ratio by County
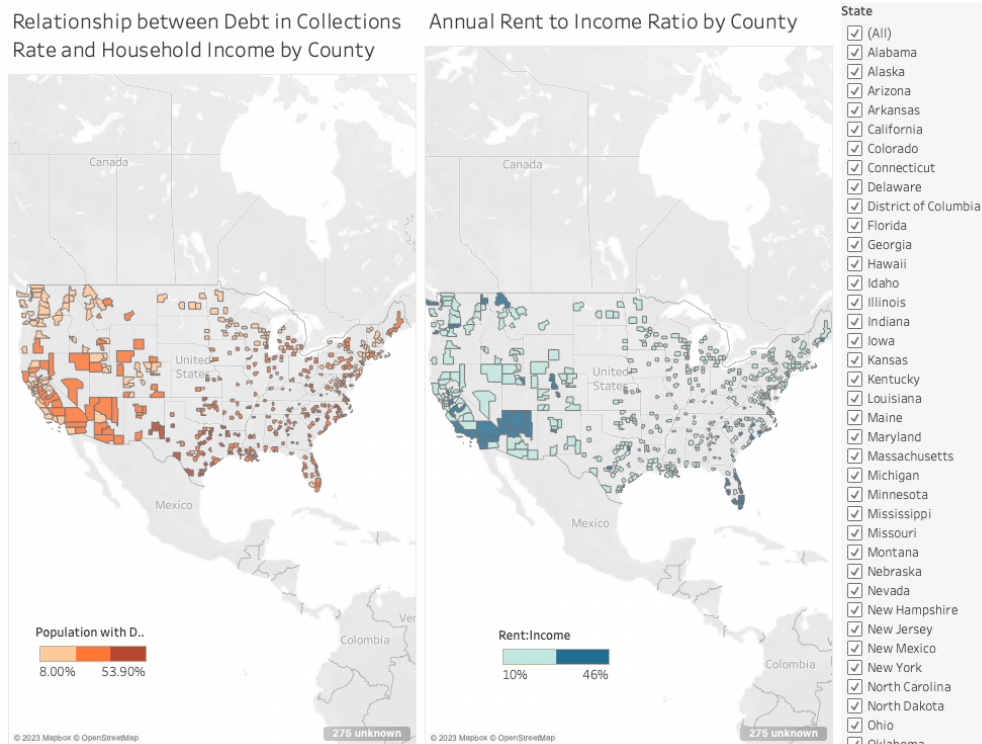
6. Conclusion

The Tableau story point visualizations present a comprehensive look into the relationship between household income, debt trends, and housing affordability across the U.S. at both state and county levels. In the state-by-state analysis, there's a juxtaposition of average household income against debt rates. Some states, despite having high average incomes, show a significant portion of the population in debt, indicating that a higher income doesn't always correlate with lower debt. Conversely those states with the highest levels of debt almost always also have the lowest income levels.

Drilling down to the county level, the impact of debt patterns on housing affordability is evident. A chart displays the correlation between the percentage of the population with debt in collections and housing costs, emphasizing the burden of rising debt on securing housing. Simultaneously, another graph highlights the connection between medical debt's share in total collections and housing affordability, suggesting that unforeseen medical expenses can have significant repercussions on housing stability.

Finally, there's a comparison of county-level median debt in collections against Zillow's estimated rent costs, showcasing disparities across regions. Simultaneously, another graph contrasts average household income with Zillow's home estimates, reinforcing the complex interplay between income, housing costs, and debt, which can vary widely across the U.S.