Final Project Essay

Allie Shafer

Rize Education

DAMII: Foundations of Data Analytics II

Dr. Darlington Sabasi

April 21, 2024

Background and Hypothesis

This analysis focuses on homeownership rates across the United States and the various features that may influence such rates. Homeownership has seemed unobtainable for many individuals, especially during the last five to ten years. By gathering insights on the features that affect homeownership rates, people looking to buy a home will better understand what to focus on. For example, suppose the distribution of homeownership rates is disproportionate across the United States. In that case, people looking to purchase homes can focus on areas with higher homeownership rates than on areas with lower rates. For this analysis, six questions are posed, with accompanying null and alternative hypotheses described in the following table.

	Question	**Type**	**Null Hypothesis**	**Alternative Hypothesis**
Hypothesis 1	How does household income and mortgage impact homeownership rates in the D.C. area?	Correlation	There is no correlation between household income, mortgage, and homeownership rates in the D.C. area.	There is a statistical correlation between household income, mortgage, and homeownership rates in the D.C. area.
Hypothesis 2	How strongly correlated are median income and average household cost	Correlation	There is no correlation between median income, average household cost, and	There is a statistical correlation between median income, average household cost, and homeownership rates.

	to the percentage of homeowners across the Midwest?		homeownership rates.	
Hypothesis 3	Does a higher homeownership rate indicate a higher number of cars per owner?	Correlation	There is no positive correlation between the number of cars per household and the homeownership rate.	There is a positive correlation between the number of cars per household and the homeownership rate.
Hypothesis 4	Does the classification of renter versus owner impact house size and the number of rooms in a house?	Causation	There is no statistical difference in house size or number of rooms between renters and owners.	There is a statistical difference in size of 200 square feet and difference of 1.8 rooms between renters and owners.
Hypothesis 5	How does region impact family size in	Causation	There is no difference between the average family size of a sample	There is a statistical difference between the average family size of a sample and the expected

	the United States?		and the expected average family size of the entire United States.	average family size from the population.
Hypothesis 6	Do the different regions of Minnesota have statistically different rates compared to the average for Minnesota?	Causation	There is no difference between the homeownership rates of the geographical regions in Minnesota and the average.	There is a statistical difference between the homeownership rates of the geographical regions in Minnesota and the average.

The results of these hypotheses will help to determine the factors prospective homeowners should consider when looking to buy or rent in a certain region of the United States.

Research Methodologies

To answer the questions described above, a number of datasets were explored. However, many of them had large quantities of missing data or simply did not contain the necessary variables. Ultimately, the Location Affordability Index from the U.S. Department of Housing and Urban Development's Office of Policy Development and Research (PD&R) and the Minnesota Homeownership Rate dataset from the Federal Reserve Bank of St. Louis were chosen. A couple of supplemental datasets were also selected to add context to the Location Affordability Index dataset. For example, the Location Affordability Index contained state and county codes, but not their names. Datasets with these codes and their respective names were joined with the Location Affordability Index to improve the ease of understanding of the data.

For each hypothesis, a standard research framework was followed. First, some initial data visualization is conducted, and an expected outcome is assessed. Second, the null and alternative hypotheses are stated. Third, the necessary variables are extracted from the dataset, and calculated if necessary. Fourth, a test statistic is determined and calculated. Fifth, the p-value or correlation value is calculated, and the null hypothesis is either supported or rejected.

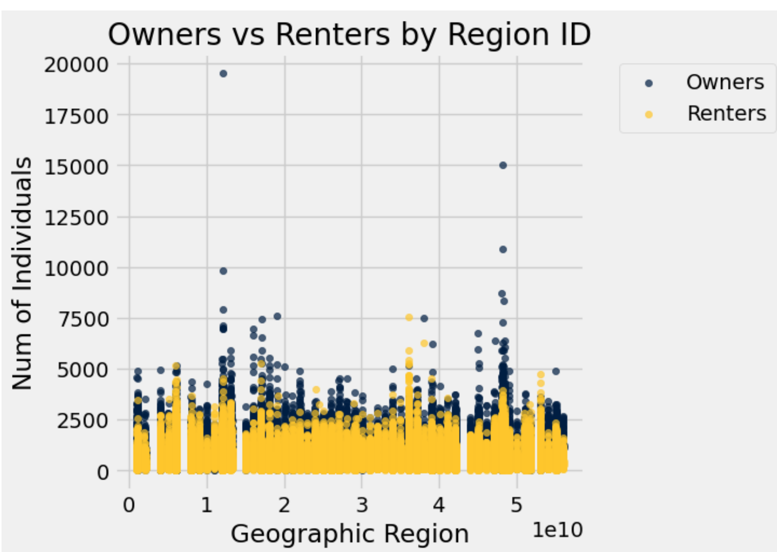
A number of variables are used in this analysis, including total households, number of owners within total households, median income, median mortgage, average household cost, median rooms per household, state name, average number of cars per owner, average family size, and the yearly homeownership rate in Minnesota. These variables are pulled from the Location Affordability Index and the Minnesota Homeownership Rate datasets. Additionally, some variables are transformed into more relevant variables, such as total households and total owners, to create a homeownership rate for each region in the Location Affordability Index dataset. Thus, the methodology for selecting data surrounds not only its usefulness for answering the above hypotheses but also in the ease of deriving additional variables as needed.

Data Collection and Sampling

Both the Location Affordability Index and Minnesota Homeownership datasets were found on open-source websites. As previously stated, the Location Affordability Index came from the Office of Policy Development and Research (PD&R), and the Minnesota Homeownership Rate dataset came from the open-source Federal Reserve Economic Data portal from the Federal Reserve Bank of St. Louis. For the Location Affordability Index, some columns and rows contained only null values, which were dropped prior to analysis. Additionally, some supplemental data had to be added, replacing state and county codes with their respective names. The entire dataset was used for questions regarding correlation, as these questions were concerned with the population dataset rather than specific regions. For questions regarding causation, sample sizes of 100 were used to ensure variability in the region. The questions regarding correlation were tested by calculating the correlation between each pair of variables and comparing it with the correlation cutoff value of 0.3 (or -0.3 if there is a negative relationship). The questions regarding causation were tested by calculating the total variation distance, or just distance in some cases, and from these differences, calculating the p-value from the expected distance. The p-value was then compared to the cutoff value of 0.05 to determine whether the null hypothesis was supported or rejected by the data.

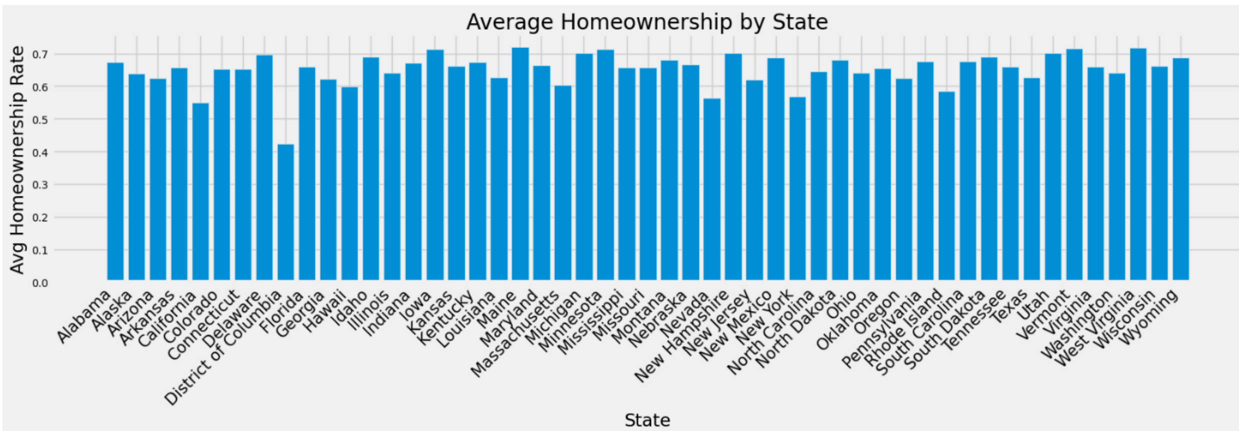
Data Overview

The section above describes the variables used for analysis. However, the relationships between these variables has yet to be explored. This section gives an overview of the anticipated

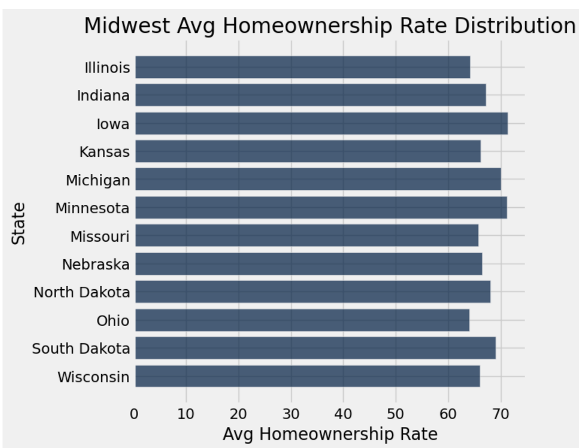


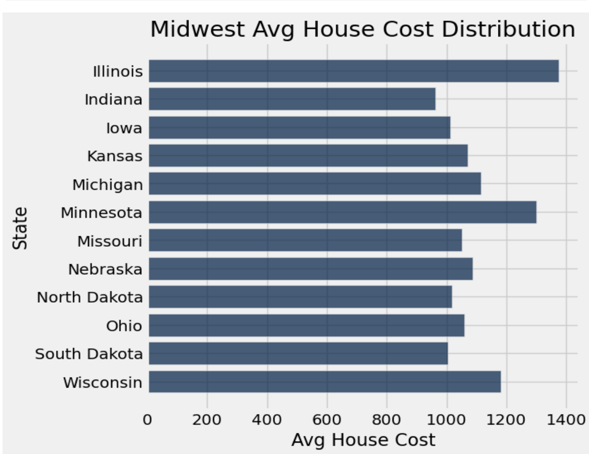
relationships between variables, as well as insights into specific statistics of the variables. The scatter plot to the left shows the number of owners and renters for a given geographic area. From this, it can be seen that there are more owners than renters. However, the amount of variability is not clear. Therefore, a homeownership rate will need to be calculated from this data.

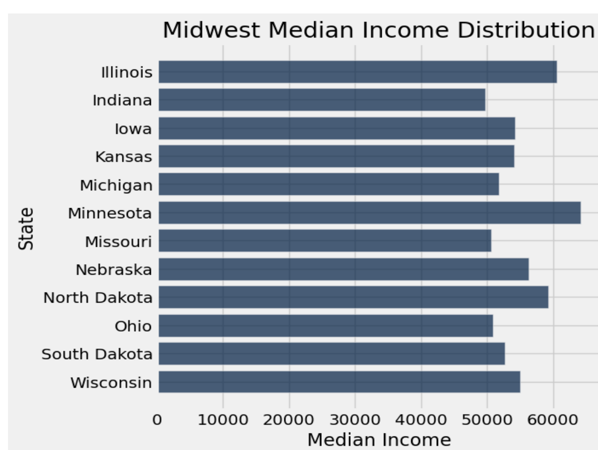
The following bar plot displays the calculated homeownership rate by state, indicating a lower homeownership rate for the D.C. area than the rest of the United States. Moreover, this bar plot showcases the need to perform a correlation analysis on the various variables of the D.C. area to identify why there is a lower-than-average homeownership rate.

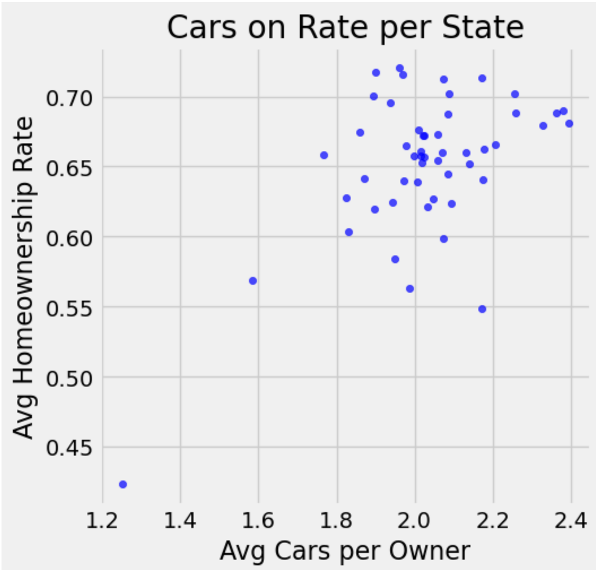


The following visualizations show the distribution of median income, average house cost, and average homeownership rate for all Midwest states. As can be seen by the similar behavior of the bars, there appears to be a relationship between median income, average house cost, and average homeownership rate. A correlation test will be used to see if this relationship is linear.



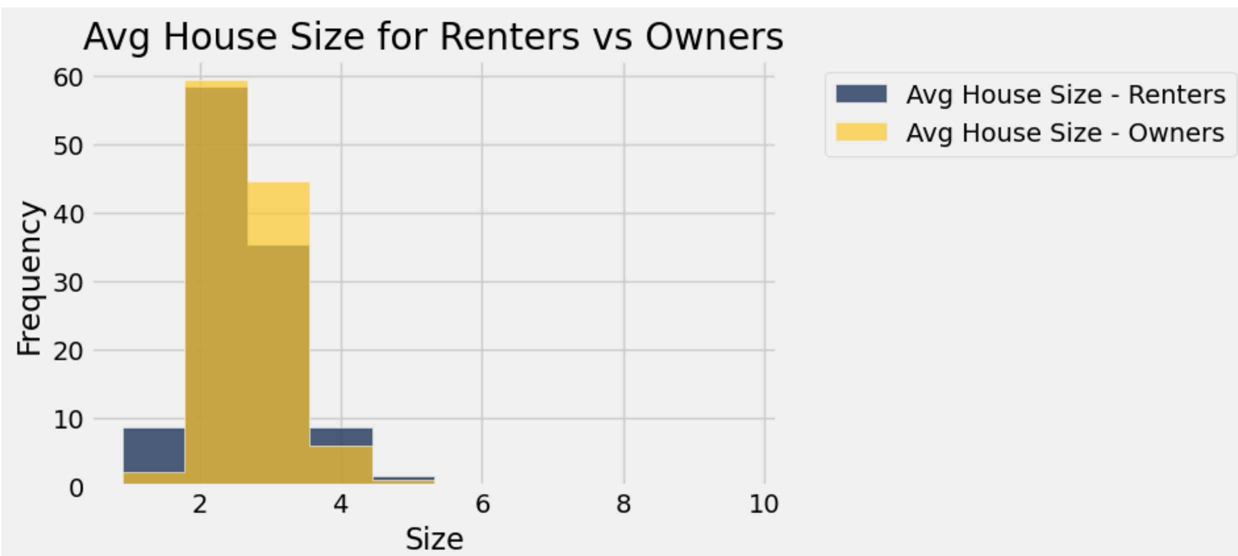




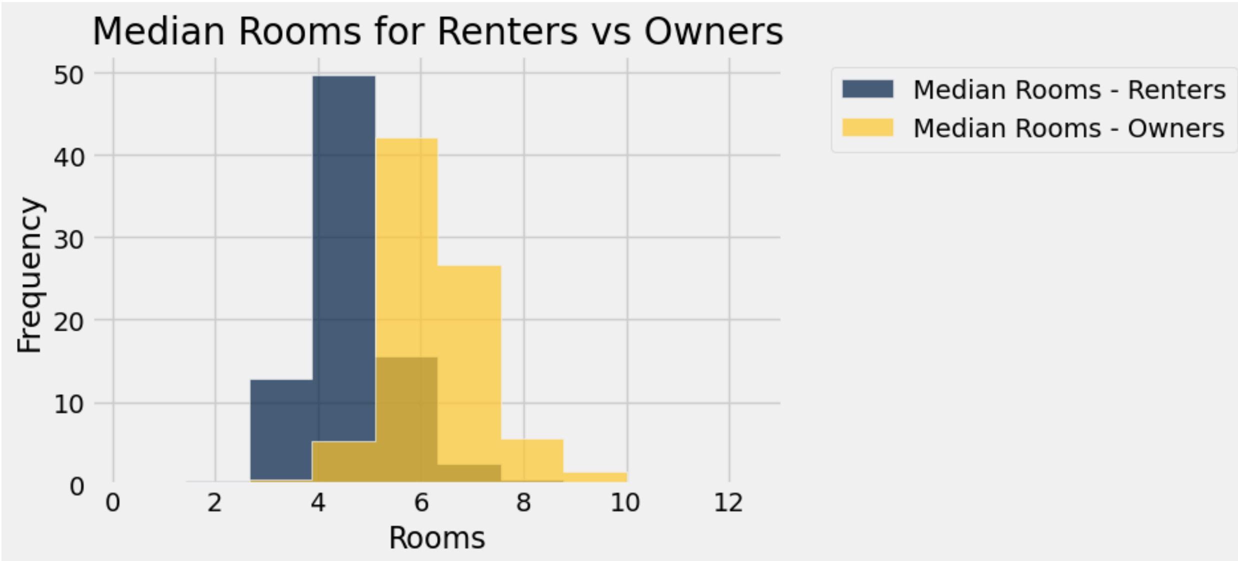


The scatter plot to the left explores the relationship between the average number of cars a household has and the homeownership of that same state. Moreover, the visualization indicates possible linearity. A correlation test and regression analysis will be conducted to explore this relationship further, and support for or against the null hypothesis will be examined.

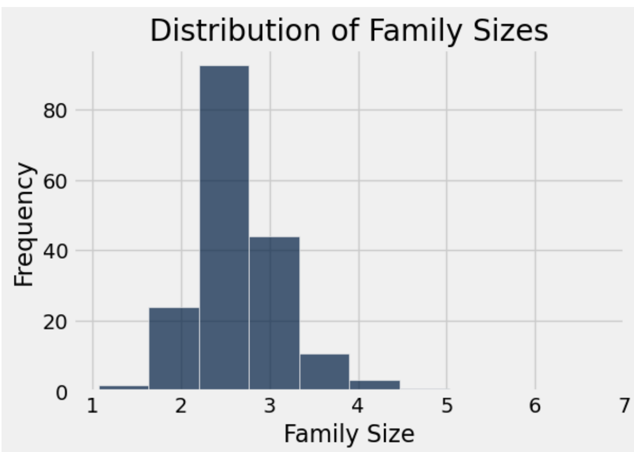
As seen in the following visualization, the average house size for renters and owners almost completely overlaps. There appears to be about a 200-square-foot difference. A hypothesis test of the differences will be conducted to accept or reject this hypothesis.



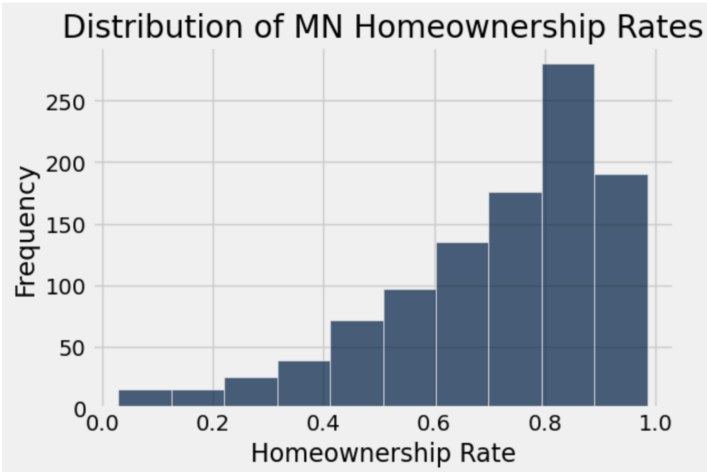
The following visualization showcases the difference in the median number of rooms between renters and owners. There appears to be a slight difference between the two, and a hypothesis test of the differences will be required to accept or reject this hypothesis.



The following visualization shows the distribution of average family sizes across the United States. As seen below, there is a wide distribution of family sizes, from 1 all the way to 7 people per family. Measuring the average family size for various regions across the United States and comparing them to the total average family size will help to identify whether the region impacts family size, thus supporting or rejecting the null hypothesis.



The following histogram shows the distribution of homeownership rates across the state of Minnesota. From the distribution, it appears that most homeownership rates center around 0.8. However, the average homeownership rate for the state of Minnesota calculated from the Minnesota Homeownership Rate dataset is 0.73. From this, a hypothesis test will be conducted using the differences between the average homeownership rates of a sample and the expected rate calculated from the Minnesota Homeownership Rate dataset. The resulting calculated p-

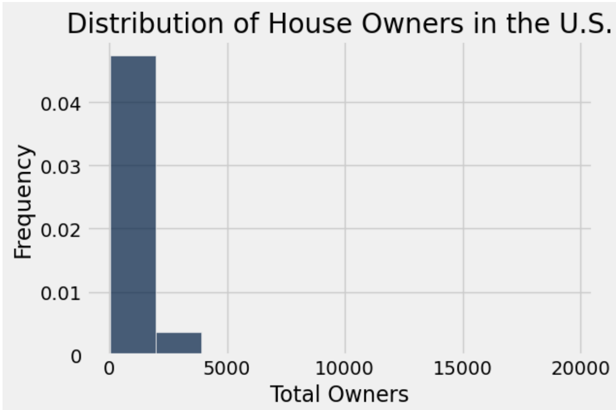


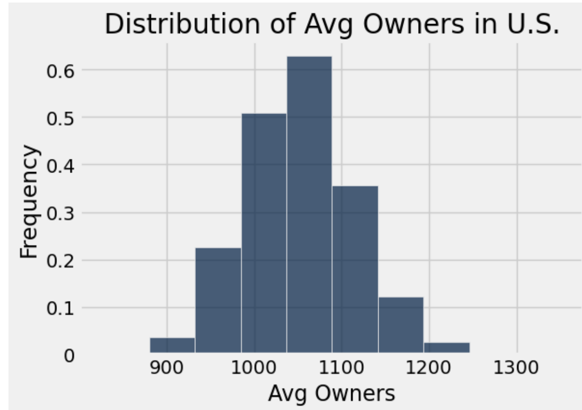
value will enable concrete conclusions to be made regarding the differences between the homeownership rates in the various regions of Minnesota. Finally, the p-value will determine whether the null hypothesis is supported or rejected for this hypothesis test.

The following table describes the measures of central tendency for each column used in this analysis. As expected, some columns have much wider ranges, such as median_hh_income, which has a range of 245380, while other columns have much smaller ranges, such as avg_hh_size, which has a range of 5.66. Moreover, the median, mean, and mode all show similar levels of variability. As a result of this variability, most values will be standardized prior to analysis to ensure that larger values do not introduce bias into the results.

Column	Range	Median	Mean	Mode
households	20431	1547	1646.61	1449
owner_occupied_hu	19517	967	1054.27	793
renter_occupied_hu	7532	475	592.341	235
median_smoc_mortgage	3623	1356	1552.96	4001
avg_h_cost	3762.99	1190.98	1351.45	1067.4
autos_per_hh_owner	3.46479	2.02473	1.98828	2
avg_hh_size	5.66	2.59	2.6503	2.48
median_hh_income	245380	52638	59015	46250
median_rooms_per_renter_hu	12.229	4.5	4.63776	4.3
median_rooms_per_owner_hu	7.6	6.2	6.27182	6.1

Finally, the following two visualizations showcase support for the central limit theorem. The distribution of the total number of owners across the United States is shown in the first histogram and is not normally distributed. By creating a histogram of the averages of samples drawn at random from the population dataset (as seen in the second visualization), the data takes on the shape of a normal distribution. This makes this data easier to understand and makes measures of central tendency much clearer to the viewer.





Overall, there are a number of patterns and relationships in this data that will be further explored in the hypothesis-testing portion of this analysis. Furthermore, measures of central tendency and the central limit theorem provide additional insights into the behavior of the underlying data, which will provide support for or against the null hypothesis.

Hypothesis Testing

Hypothesis 1

For the first hypothesis, data is extracted and explored to try and understand the correlation between household income, mortgage, and the homeownership rate. Furthermore, this question was broken down into three hypothesis tests for the District of Columbia (D.C.) area: one tested the correlation between household income and the homeownership rate, one tested the correlation between the mortgage amount and the homeownership rate, and one tested the correlation between household income and mortgage amount. The null and alternative hypotheses for the correlation between income and homeownership rate are as follows:

- Null Hypothesis: There is no correlation between the median household income and the homeownership rate within the D.C. area.

- Alternative Hypothesis: There is a correlation between the median household income and the homeownership rate within the D.C. area.

The null and alternative hypotheses for the correlation between mortgage and homeownership rates are as follows:

- Null Hypothesis: There is no correlation between the median mortgage and the homeownership rate within the D.C. area.
- Alternative Hypothesis: There is a correlation between the median mortgage and the homeownership rate within the D.C. area.

The null and alternative hypotheses for the correlation between mortgage and income are as follows:

- Null Hypothesis: There is no correlation between the median household income and the median mortgage of the D.C. area.
- Alternative Hypothesis: There is a correlation between the median household income and the median mortgage of the D.C. area.

The test statistic used for both hypotheses was the correlation value calculated between each pair of values. A cutoff value of 0.3 (or -0.3 for negative correlations) was used to reject the null hypothesis. For the first hypothesis, a correlation of 0.61 was calculated. For the second hypothesis, a correlation of 0.39 was calculated. For the third hypothesis, a correlation of 0.85 was calculated. From these correlations, a regression analysis was performed, and regression models for each pair of variables were developed. From this correlation and regression testing, additional insights were gleaned, which helped to answer the question of what influences homeownership rates in the D.C. area. This additional analysis and the insights gathered will be explored in more detail in the section titled “Interpreting the Results.”

Hypothesis 2

Much like in the previous hypothesis test, two sets of null and alternative hypotheses are used to test for correlation. Moreover, correlation is tested to identify how strongly correlated median income and average household cost are to the percent of homeowners (homeownership rate) across the Midwest. The null and alternative hypotheses for median income and homeownership rate are as follows:

- Null Hypothesis: There is no correlation between median income and the homeownership rate in Midwest states.
- Alternative Hypothesis: There is a statistical correlation between median income and the homeownership rate in Midwest states.

Hypotheses for average household cost on homeownership rates are as follows:

- Null Hypothesis: There is no correlation between average household cost and the homeownership rate in Midwest states.
- Alternative Hypothesis: There is a statistical correlation between average household cost and the homeownership rate in Midwest states.

The test statistic that was used to test these hypotheses was the correlation between each pair of values. The correlation cutoff value was 0.3 (or -0.3 for a negative correlation). For this hypothesis, only a correlation test was conducted. The correlation between income and homeownership rate in the Midwest was calculated to be 0.18, and the correlation between average house cost and homeownership was calculated to be 0.04. Moreover, the meaning of these correlations is described in the section titled "Interpreting the Results."

Hypothesis 3

For this hypothesis, the correlation between the number of cars and the homeownership rate is determined. The null and alternative hypotheses are described below:

- Null Hypothesis: There is no positive correlation between the number of cars per household and the homeownership rate.
- Alternative Hypothesis: There is a positive correlation between the number of cars per household and the homeownership rate.

As with the previous two hypotheses, the test statistic used was the correlation between the number of cars and the homeownership rate. The correlation cutoff value will be 0.3 only, because this hypothesis is specifically testing for positive correlation. A regression analysis was also performed, with an analysis of the residuals providing insight into the model created. The correlation between the number of cars per household and the homeownership rate was calculated to be 0.57. Moreover, the results of the correlation test and regression analysis are described in greater detail in the section "Interpreting the Results."

Hypothesis 4

The fourth hypothesis seeks to understand if there is a causal relationship between the classification of owner or renter with the average house size and average number of rooms. It is expected that renters will have smaller houses with a smaller number of rooms. The null and alternative hypotheses for the relationship between owners, renters, and the average house size is as follows:

- Null Hypothesis: There is no statistical difference in house size between renters and owners.
- Alternative Hypothesis: There is a difference in size of about 200 square feet (0.2) or larger between houses selected by renters and houses selected by owners.

The null and alternative hypotheses for the relationship between owners, renters, and the number of rooms are described below:

- Null Hypothesis: There is no difference in the median number of rooms between houses being rented versus houses that are owned.
- Alternative Hypothesis: There is a statistical difference of 1.8 or higher between the median number of rooms in houses that are rented and houses that are owned.

For both sets of hypotheses, the test statistic is the absolute value of the differences between the variables of renters and owners. The p-value for the hypothesis regarding house sizes between renters and owners is 3%. The p-value for the hypothesis regarding number of rooms between renters and owners is 3.8%. The meaning of the p-value will be described in the section titled "Interpreting the Results."

Hypothesis 5

This hypothesis is focused on understanding the variability in family size across the United States. The United States offers tax breaks to those who are married, and for each child they have. Therefore, regions that are more expensive to live in might have family sizes that are larger in order to receive tax breaks. This hypothesis test is not concerned with the explanation behind family size but rather is looking to seek answers on whether average family size varies across regions in the United States or if it tends to follow the population average. The central limit theorem states that averages from samples tend to be normally distributed, and it is expected that the averages center around the population average. According to Statista, a difference between the average family size of a sample and the average family size of the population of 0.1 or larger would indicate that there are regions with much higher concentrations of family sizes that are far away from the population mean (Statista, 2024). Thus, the null and alternative hypotheses can be described as follows:

- Null Hypothesis: There is no difference between the average family size of a sample and the expected family size.
- Alternative Hypothesis: There is a statistical difference between the average family size of a sample and the expected family size from the population of 0.1 or larger.

The test statistic that was used for this hypothesis test is the absolute value of the difference between the average family size of a sample and the average family size of the Location Affordability Index dataset. Using the expected difference of 0.1, the p-value was calculated to be 4.8%. The significance of such a value will be described in the section titled "Interpreting the Results."

Hypothesis 6

The last hypothesis of this analysis is concerned with the causal relationship between the regions of Minnesota and the total average homeownership rate in that state. The Minnesota Homeownership Rate dataset was used to calculate the average homeownership rate between 2012 and 2016, which are the same years covered in the Location Affordability Index. This value forms the expected average. From there, the following null and alternative hypotheses can be declared:

- Null Hypothesis: There is no difference between the homeownership rates of the geographical regions in Minnesota.
- Alternative Hypothesis: There is a statistical difference between the homeownership rates of the geographical regions in Minnesota.

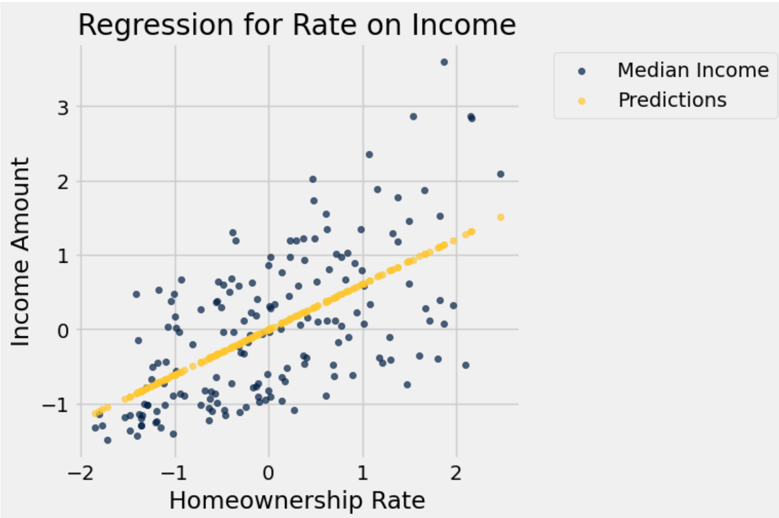
The test statistic used to test this hypothesis is the absolute value of the difference between the average homeownership rate of a sample and the expected state homeownership rate. Large differences indicate that the homeownership rate in Minnesota is disproportionately distributed across the various regions from which samples were drawn. The p-value for this test was calculated to be 5.3%. The significance of such a value will be explored in the section title "Interpreting the Results."

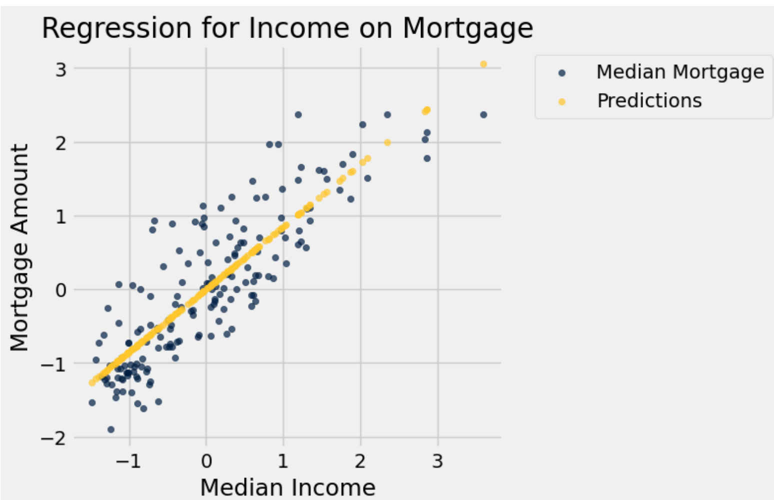
Interpreting the Results

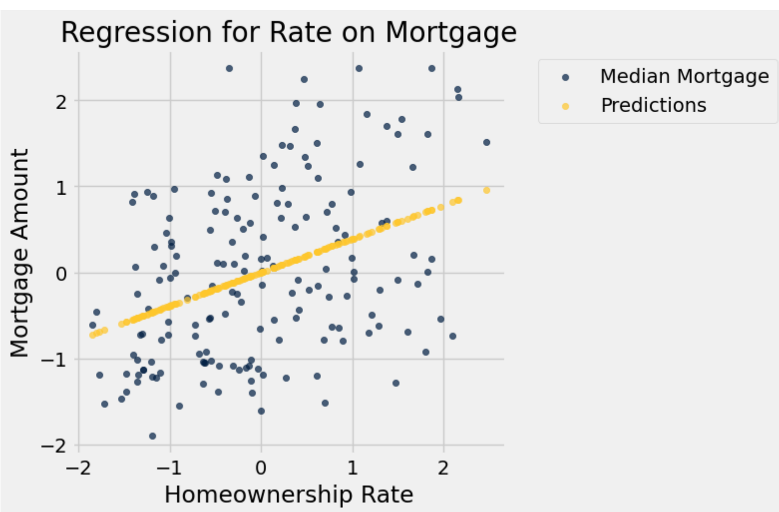
Hypothesis 1

As previously stated, the correlation between income/mortgage and income/homeownership is much higher than the correlation between mortgage/homeownership for the D.C. area. In terms of the stated hypotheses, each calculated correlation is larger than 0.3, so all null hypotheses will be rejected. In other words, there is a statistical correlation between median income and median mortgage, median income and homeownership rates, and median mortgage and homeownership rates. Moreover, all three correlations are positive, indicating a positive linear association. Thus, it can be concluded that a higher homeownership rate is associated with higher median incomes and mortgages. Moreover, as incomes increase, there is a positive change in mortgages as well.

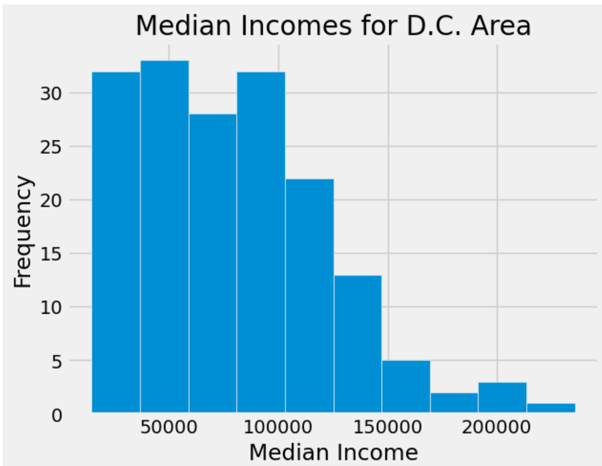
While the individual correlation values are very helpful, it is also important to review the actual regression line against the data. This will help to identify which regression lines are the best fit for the data and, therefore, which features are best correlated. The following three visualizations show the regression models created for each set of variables. These visualizations show that the homeownership rate is a good indicator of income and mortgage. Due to the positive linearity, income and mortgage are also good indicators of homeownership rates.

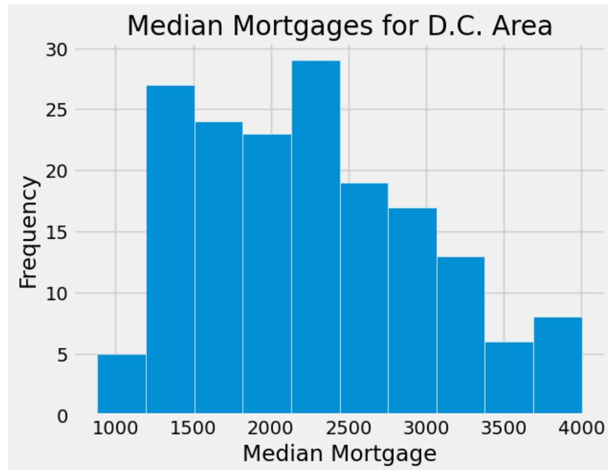






The homeownership rate within the D.C. area was much lower than average, with a rate of around 0.45 as opposed to the average of 0.65. Thus, the regression analysis conducted to the left gives a great deal of insights into homeownership rates in the D.C. area. First, there is a statistical correlation between mortgages and incomes, incomes and homeownership rates, and mortgages and homeownership rates. Second, this means that incomes, mortgages, and homeownership rates can be used as predictors, as seen by the linear regression models calculated for each feature. Therefore, it would not be unreasonable to assume that a wide distribution of incomes and mortgages would result in a wide distribution of homeownership rates. Moreover, the distributions of incomes and mortgages for the D.C. area could help to solidify an answer regarding both the lower-than-average homeownership rate for D.C., as well as the extreme variability in homeownership rates within the area. These distributions can be seen below.



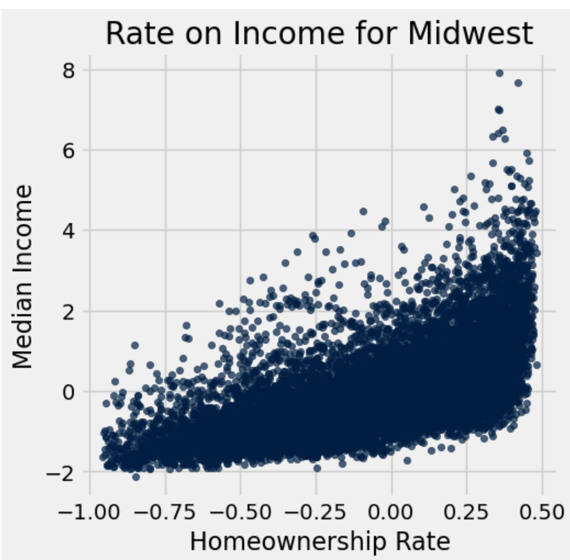


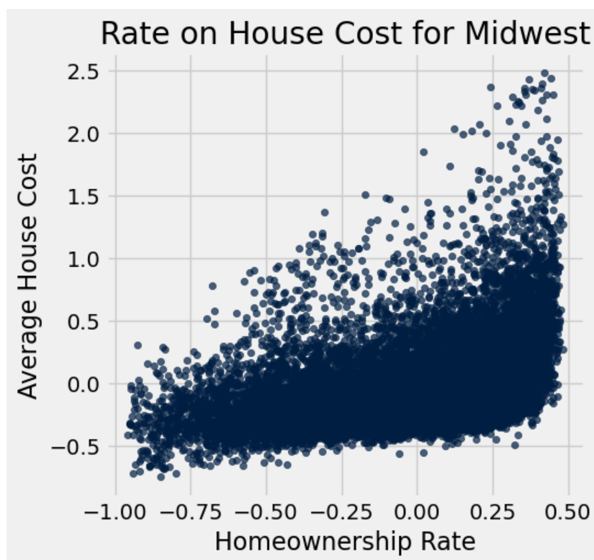
As seen in the above visualizations, the median incomes are skewed to the right, which means there is a higher concentration of low median incomes with a few very high median incomes. However, the median mortgages are more uniform in nature, centering between 2000 and 2500. While there is some very slight right skewness occurring, it is nowhere near the level of skewness seen in the median income histogram. From these histograms, it can be concluded that the mortgage amounts available are disproportionate to the distribution of median incomes, thereby resulting in a much lower homeownership rate compared to the rest of the U.S. Finally, the wide distribution of median incomes helps to explain the variability in homeownership rates for individual areas within D.C. A concentration of lower-income individuals would result in a lower homeownership rate, whereas a concentration of higher-income individuals would result in a higher homeownership rate. Therefore, it can be concluded that the distribution of incomes is not even across the D.C. area, resulting in some areas with exceptionally high homeownership rates and other areas with exceptionally low homeownership rates.

Overall, the hypothesis tests conducted conclude that there is a positive linear relationship between income, mortgage, and homeownership rates within the D.C. area. From these conclusions, the distributions of income and mortgage showcase the disparity resulting in a lower than average homeownership rate when compared to the rest of the United States.

Hypothesis 2

As previously stated, the correlations for income/homeownership rate and cost/homeownership rate are 0.18 and 0.04, respectively. This means that the correlation between income/homeownership and house cost/homeownership is very low. With both correlations below 0.3, the null hypothesis is supported. Therefore, there is no linear correlation between median income and homeownership rates or average house cost and homeownership rates. This is further supported by a visualization of the variables, as seen below. A lack of linearity is also why a linear regression analysis was not performed.



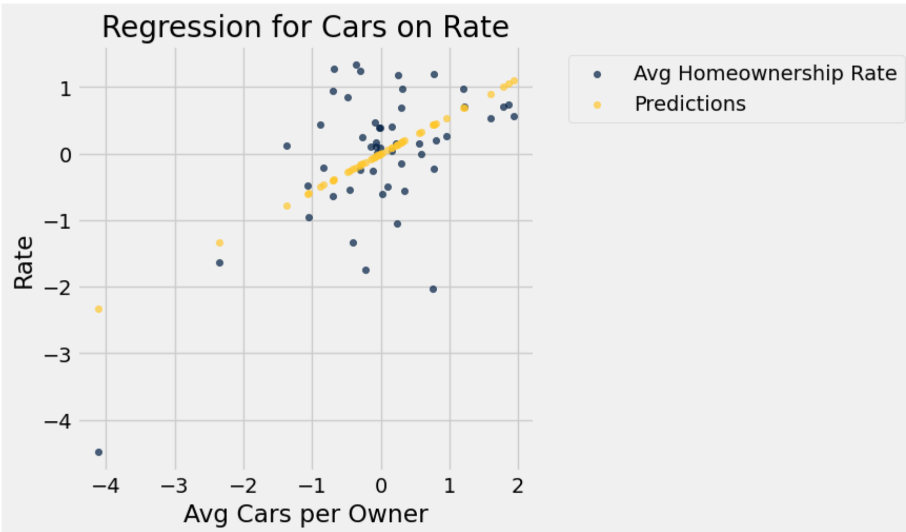


From the above scatter plots, there appears to be an exponential relationship between the variables. However, confirming or denying that claim is outside the scope of this hypothesis.

Thus, all that can be definitively confirmed is that there is no linear correlation between the variables.

Hypothesis 3

With a correlation of 0.57, the null hypothesis is rejected. This means there is a positive correlation between the number of cars per owner and the homeownership rate in the United



States. A regression analysis can be performed to identify the regression line. The visualization to the left shows the results of the regression analysis.

As expected, the correlation between the number of cars per owner and the homeownership rate is positively correlated. From this, it can be concluded that a family with a higher number of cars is more likely to also have a house than those who have a smaller number of cars. It is important to note, however, that just because someone has multiple cars does not mean they will automatically have a house. Moreover, a high homeownership rate for a certain area does not necessarily mean the families within those houses have multiple cars. The correlation statistic simply means that the behavior of homeownership rates and cars per family increases together, which is further supported by the calculated regression line.

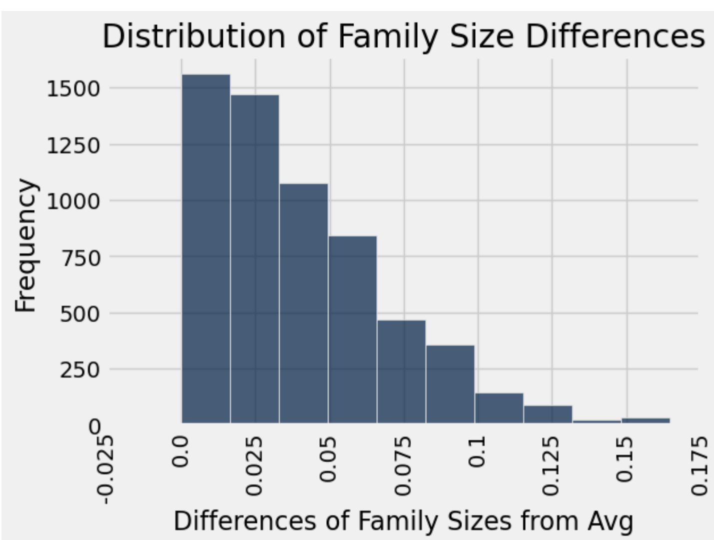
Hypothesis 4

As previously stated, the p-value for the hypothesis testing causation between renters, owners, and house size is 3%. With a cutoff value of 5%, the calculated p-value rejects the null hypothesis. This means there is a difference of about 200 square feet or greater between the houses selected by renters and the houses selected by owners.

According to iProperty Management, the difference in the average number of rooms between renters and owners is expected to be 1.8 (iProperty Management, 2023). Under the null hypothesis, this difference is expected to be zero. However, the p-value is 3.8%, which is below the cutoff value of 5%. This means the null hypothesis can be rejected, and the differences between renters and owners regarding the median number of rooms in a house are statistically significant. In other words, there does exist a difference between the median number of rooms in a renter's house compared with an owner's house in the same area by 1.8 or higher.

Hypothesis 5

From the p-value, it can be concluded that there is a statistical significance between the



sample average family sizes and the expected average family size. The p-value is 4.8%, which falls below the cut-off value of 5%, indicating statistical significance. This means average family sizes tend to vary quite a bit across the United States, differing by more than 0.1, which was the expected difference used for hypothesis testing and calculating the p-value. A distribution of the differences in family sizes from the average family size can be reviewed above.

Hypothesis 6

It is expected that the average homeownership rate would differ by about 0.5% across the state of Minnesota (Eckes, 2024). Anything higher than this would indicate an uneven distribution of homeownership rates across the various regions that make up Minnesota. With a p-value of 5.3% and a p-value cutoff of only 5%, it can be concluded that the null hypothesis is supported by the data for this hypothesis. This means there is no statistical significance in the average homeownership rates for the various geographic regions of Minnesota and the average homeownership rate of Minnesota. Moreover, the homeownership rate in Minnesota tends to be steady across the state and is not centralized in a specific geographic region. In other words, homeownership is accessible at the same rate across the state of Minnesota.

Interpretation of Analysis

While some of the p-values and correlation values rejected the null hypothesis for various hypothesis tests, that does not mean the results apply in the real world. The Location Affordability Index retains data from the 2012 to 2016 period, which was almost 8 years ago at the time of performing this analysis. Thus, the patterns discovered in this dataset may not hold true in today's economic climate. Fortunately, this dataset was incredibly large, with almost 70000 rows of data to work with. As a result, any decision errors due to sampling size were mitigated by the extensive amount of data that was being drawn from. Furthermore, possible decision errors were also mitigated by using a small p-value cutoff value of 5%, which ensures that 95% of the time (or more), the conclusion of the hypothesis test is correct. Finally, a frequentist approach was taken in this analysis. This means samples were repeatedly drawn from observed population data in order to make a claim about a variable in the population. This approach was best suited for this analysis, as the Location Affordability Index was created out of observed values, and the large quantity of data available enabled numerous testing or trials to be performed.

Conclusion

Overall, the Location Affordability Index and Minnesota Homeownership Rate datasets provide ample data for a variety of hypothesis tests to be performed. Together, the conclusions of such tests provide greater insights into the variables affecting homeownership rates in various regions across the United States. Moreover, the results of this analysis provide a starting point for individuals looking for ways to identify areas with affordable homes. For example, the correlation between the average number of cars and the homeownership rate indicates that those with houses also tend to have more cars than those who just rent. This does not mean that having a higher number of cars will automatically result in homes being more accessible. Rather, it is an indication that houses will most likely be built with additional space for more cars, potentially making them more expensive and, therefore, unobtainable to some. Many individuals search for houses with fewer bedrooms to lower costs. However, if an individual is looking to purchase a house that is in line with a lower budget, searching for houses with fewer or smaller garages would also be a good place to start. Another example concerns the correlation between median income, median mortgage, and homeownership rates in the D.C. area. As a result of the regression analysis conducted with these variables, an individual who knows of the homeownership rate in a D.C. area can utilize the regression models in order to estimate the incomes and mortgages of that area. By comparing those numbers with the individual's current income and/or mortgage, they will have a good indication of their own success in that area. Thus, the results of the hypothesis tests, and whether they support or reject the null hypothesis, serve as an assessment of the variables impacting homeownership rates across the United States, which enables greater decision-making on the part of the homebuyer.

References

Eckes, R. (2024, Feb 2). *Examining Minnesota's current homeownership gap*. Twin Cities Habitat for Humanity. https://www.tchabitat.org/blog/examining-minnesotas-current-homeownership-gap#:~:text=Their%202015%2D2020%20study%20shows,Black%20homeownership%20was%20the%20lowest

Federal Reserve Bank of St. Louis. (2023). *Homeownership rate for Minnesota*. [Data set]. FRED. https://fred.stlouisfed.org/graph/?id=MNHOWN

iProperty Management. (2023, Nov 22). *Homeowners vs renters statistics*. https://ipropertymanagement.com/research/renters-vs-homeowners-statistics

Statista. (2024). *Average number of people per family in the United States from 1960 to 2022*. https://www.statista.com/statistics/183657/average-size-of-a-family-in-the-us/

USDHUD-PD&R. (2023, Jul 31). *Location affordability index* (Version 3). [Data set]. https://hudgis-hud.opendata.arcgis.com/datasets/HUD::location-affordability-index-v-3/about