Home Ownership Rates in the United States Between Generations

Kyle Vavasour

Introduction

In recent years there has been an increasing pessimism about the prospect of owning a home, with younger generations dissatisfied how older generations own homes while they cannot afford to. Younger generations may question whether they will ever be able to own their own homes, and compare themselves to the seemingly far favorable conditions for older generations; indeed, home ownership and housing affordability is often cited as a core issue for younger generations [3].

The goal of this project is to examine differences in home ownership between generations through home ownership rates by year, age, and locations. The proposed research question is: Have home ownership rates in the United States changed between the generations of Baby Boomers, Gen X, Millennials, and Gen Z?

Answering this question is not as simple as comparing home ownership rates directly between generations; as we might expect, generations that have been alive longer will tend to have higher rates due to more time to gain wealth and more time in the housing market. Instead, to better compare home ownership rates, we have to look at the trends between generations of time, age, and geography to see if there are real differences. Younger generations cite home ownership as a core issue, but to get a sense of whether home ownership has truly declined, an analysis of the trends over age, time, and geography must be preformed. Conducting this analysis will provide useful insight into the issue of home ownership between generations.

The dataset that was used to answer this question is a subset of the IPSUM CPS dataset. The dataset comes from the United States Census Current Population Survey (CPS) and its supplemental Annual Social and Economic Supplement survey (ASEC). The CPS is preformed monthly to different households in the US while the ASEC supplement is preformed annualy in March. The dataset contains both household and individual level information, meaning some variables refer to entire household information such as whether a house is owned or rented, while other variables refer to individual information such as a persons age.

As the entire IPUMS CPS dataset is large and contains many variables that are irrelevant to this analysis, a subset of the data was used containing only the variables and year range of interest described in the methods section.

Methods

The methods section includes subsections corresponding to how the data was obtained, data cleaning, data wrangling, methods for producing plots, and methods for predictive modelling.

Data Collection

IPUMS CPS allows users to create their own subsets, called data extracts, of the full CPS dataset that contain variables and year ranges of interest. To create a data extract, a user must select variables and year ranges and then wait for a data extract to be generated; once the extract is generated it can be downloaded. The creation and retrieval of the data extract used in this analysis was preformed through the IPUMS API. To create the extract through the API, a post request was made using the httr R package with a formatted JSON string that specified the year range and variables of interest. The year range was 1976-2023 and the variables of interest were:

-YEAR which is the year an observation was from

-AGE which is the age of an individual

-OWNERSHIP which is whether the household that an individual belongs to is rented/owned

-RELATE which is the relation to their relation to the householder of the household an individual belongs to

-STATEFIP which is the FIP code of the state were the observation was from

-COUNTY which is the FIP code of the county were the observation was from

-HHINCOME which is the total household income of the household an individual belongs to

Once the data extract request was submitted, a get request was used to check whether the data extract had been generated. Finally, when the data extract was complete, the data files were downloaded from the URLs provided from the content of the get request. The dataset was then loaded in R using the ipumsr package provided by IPUMS for loading IPUMS data into R.

In addition to the key variables mentioned in the introduction, IPUMS provides weights for calculating individual and household level statistics to mitigate the sampling bias of the the survey, these weights were also included in the dataset as the variables ASECWTH for household level statistics and ASECWT for individual level statistics.

Finally, more data was acquired for use in the predictive modelling section. Historical average mortgage interest rates in the US between 1976-2023 were obtained from Federal Reserve Economic Data webiste [4]. Additionally, data on historical median home prices in the US between 1976-2023 was obtained from the DQYDJ website [5]. Finally, data for adjusting IPUMS economic variables such as HHINCOME for inflation was obtained from the IPUMS website which provided conversion rates for each year to standardize the variables to the year 1999.

Data Cleaning

Once the dataset was loaded into R, it's dimensions were checked for how many observations were present in the dataset and whether the number of variables was consistent with what was requested in the data extract. Checking the dimension showed that the dataset contains 8,293,750 observations with each observation being an individual who belongs to a household. The head, tail, and variable types were also checked to make sure the dataset was imported correctly.

Once the dataset was confirmed to have been imported correctly, the variables were checked for missing values and suspicious values. All values for AGE, YEAR, RELATE, OWNERSHP, STATEFIP, HHINCOME, and INCRETIR had no missing values and were all in their expected range; however, the variable COUNTY had 3,159,092 missing values and 2,991,865 values of 0 which is not a recognized county FIP code. Checking the IPUMS data dictionary confirmed that a value of 0 was used for counties that did not meet a population threshold due to concerns about deanonymizing data. Further investigation also showed that all the missing values in the COUNTY variable occurred before 1995. Removing observations with missing values for COUNTY would remove a significant amount of the data. Additionally, the COUNTY variable was not used across most of the analysis; therefore, no observations were removed.

Something to note in the dataset is that the values in the categorical variables are given by codes that correspond to their meaning, for example, the variable OWNERSHIP has a value of "10" when the observation belongs to a household that is owned. The meaning of these codes is easy to find on the IPUMS CPS website, so these values were not changed in this analysis. The additional data aqquired on median home prices was already adjusted for inflation, so no further adjustments were required.

Data Wrangling

The number of observations with age >90 were small which created outliers in the home ownership rates within these age ranges. Additionally, In the United States, people cannot own property until they reach the age of majority which is 19 in some states. Therefore, the age ranges of <19 and >90 were removed from the data set. After removing all observations with age less than 19 and greater than 90 there were 5,796,909 observations left. Removing these entries means that the population of interest in this analysis are adults in the United States between the ages of 19 and 90.

For the analysis, a new variable had to be created referring the generation that individual belongs to. There are no official age cutoffs for what defines a generation, so instead the generation age ranges were chosen to be consistent with an article by Pew Research Center, the age ranges were: 1946-1964 (Baby Boomer), 1965-1980 (Gen X), 1981-1996 (Millennial), 1997-2012 (Gen Z) [2]. The generation variable was created by subtracting the AGE variable from the YEAR variable to get a birth year and then checking which generation range it was in. Finally, the frequencies of each generation was checked to see if the variable was create properly, as you might expect, Gen Z had the least amount of observations at 86,029 while Baby Boomers had the most 2,147,536 due to the data set containing many years before any individual in Gen Z was born.

For the predictive modelling, HHINCOME had to be adjusted for inflation so comparisons between observations would be accurate. To do this, the inflation conversion data acquired from IPUMS was used; the data provided a conversion rate for each year from 1976-2023 that could be used to standardize the income data to dollars in the year 1999. Using these conversion rates, the variables HHINCOME and INCRETIR were multiplied by the rate corresponding to the year that the observation was collected in.

Methods for Plots

Once the data cleaning and wrangling was complete, the analysis could be preformed. To answer the research question, home ownership rates had to be calculated across generations. For this analysis, a home ownership rate was defined:

$$\text{Home Ownership Rate (\%)} = 100 \cdot \frac{\text{Owner Occupied Households}}{\text{Total Occupied Households}}$$

This definition is consistent with the definition given by the Census Housing Vacancy Survey [2]. As the dataset provides weights for calculating household level statistics, the home ownership rate was calculated as the the sum of the weights associated with an owner occupied household divided by the sum of weights associated with an occupied household. Finally, a generations home ownership rate is calculated the same, except only households owned/occupied by an individual in that generation are counted.

To find owner occupied households within the dataset both the OWNERSHP and RELATE variable had to be used. The OWNERSHP variable in the dataset refers to whether the household the observation belongs to is owned or rented; this does not mean that the individual is a homeowners as, for example, an individual living with parents who own a home would have a value of ownership rather than rented in the OWNERSHIP variable. Instead, both the variables OWNERSHP and RELATE have to be used to find homeowners. If OWNERSHP indicates that the observation belong to a household that owns their home and RELATE indicates that the observation is the householder then that observation is considered a homeowner.

Finally, visualizations and summaries were created to explore the differences in homeownership rates between generations.

First, table summaries were created to show homeownership rates by generation in different years and at different ages. To create the table of homeownership rate by year between generations, homeownership rates were calculated for each year grouped by generation. To create the table of homeownership rate by age between generations, homeownership rates were calculated for each age grouped by generation; for example, the homeownership rate for 25 year old Baby Boomers would be calculated using only observations were the age was 25 and the observation was a Baby Boomer (these would be from years between 1971-1989).

To further explore the trend of homeownership rate over time, a line chart was created to show homeownership by year with a trend line for each generation. Similarly, to explore the trend of homeownership by age, a line chart was created to show homeownership by age with a trend line for each generation.

To explore difference in homeownership in locations by generation, a map visualization was created to show home ownership at the state level, and at the county level. To create these map visualizations, geojson files of USA state and county maps that included FIPS data were used. Home ownership rates were calculated by state/county and then this data was merged into the geojson files using the FIPS codes. For the county map, some counties were not identified, these counties were merged together within state borders to form a map that contained homeownerhip rates in all identified counties, and home ownership rates in states not including identified counties within the state. Similarly, any county with <50 observations in any generation was combined with the non-identified counties within state borders.

Methods for Modelling

The goal of this section is to model home ownership rates over time between generations, ideally creating a model capable of predicting home ownership trends into the near future. This is a complicated task as home ownership rates likely depend on a large amount of external factors such as housing market conditions, economic conditions, and social factors. While the IPUMS CPS data set provides some measures that may be useful in prediction, such as household income and retirement income, the variables available are limited. To compensate for some missing variables, data for mortgage interest rates and median home prices by year was merged into the data set.

There are also a variety of ways to approach modelling home ownership by generation. One possibility is to model home ownership at the household level; this would involve creating a model that predicts whether a specific household is owned or not (or probability that a specific household is owned) using household level data such as the homeowners income and home prices within the area. An advantage to this approach would be the ability to incorporate specific data such as geography and age of homeowner. Another approach is to model home ownership rates directly, predicting population level home ownership for a given year between generations based on population level information such as 30 year fixed interest rates for mortgages and mean income for people within the generation. An advantage to this approach is that it directly models the outcome, that being home ownership rates for a given year and generation, making it more effective at predicting future home ownership rates. For this analysis, the second approach was chosen due to the model using relatively simpler population level data compared to the more complex household level information making it easier to apply in predicting future trends.

An initial GAM was fit using home ownership rate as the outcome with year as a smooth term and generation as a factor for smooth interactions; this results in separate smooths fit for each generation that share a smoothing parameter. For the smooth term, 8 basis functions were used and a second derivative penalty was also used. A plot for the GAM was produced and compared to the observed trend to assess the fit of the model. Next, another GAM was fit using the same parameters but withholding the last 5 years of data within the data set to test how well the model extrapolates to trends five years in the future.

Next, a second GAM was fit using home ownership rate as the outcome with average household income and average retirement income as smooth terms with generation as a factor for smooth interactions as well as 30-year fixed mortgage rates and median home price as smooth terms without interaction. For the income smooth term, 3 basis functions were used and a first and second derivative penalty were used. For the home price and mortgage rate smooth terms, 10 basis functions were used with a second derivative penalty. A plot for the GAM was produced and compared to the observed trend to assess the fit of the model. Next, another GAM was fit using the same parameters but withholding the last 5 years of data within the data set to test how well the model extrapolated trends five years in the future.

Summaries of the models where produced to assess the significance of the terms and obtain deviance explained and r-squared values. For the models with withheld data, mean absolute error was calculated within the year ranges that were withheld. The five year projections of both GAMs trained with withheld years were compared against the true five year trend to assess reliability and accuracy.

Analysis

The analysis section includes subsections corresponding to the initial analysis subsection where plots where produced and analyzed and the modelling subsection where a predictive model was created and assessed.

Initial Analysis

Table 1: Home Ownership Rates (%) between 2018-2023 by Generation

Year	Baby Boomer	Gen X	Millenial	Gen Z
2018	76.94919	66.50245	40.61062	21.25693
2019	77.22946	66.92789	43.23064	22.49573
2020	78.74693	69.01683	47.94009	22.56327
2021	78.47346	69.02732	48.67208	23.36040
2022	77.82291	69.72024	51.45640	25.71410
2023	77.99646	71.36857	54.23662	25.87277

This table shows the homeownership rates by generation in recent years. Since the older generations like Baby Boomers and Millennials have been alive longer, and likely have more wealth, we would expect to see that their home ownership rate is much higher as shown in the table. We also see that the home ownership rate for Millennials has increased significantly in this time period, around a 14% difference with a signicant increase in 2020. Most Millennials would be in their late 20s and early 30s which is around the age of the average first time homebuyer in the United States. There were also historically low interest rates between 2019-2021 which could have contributed to this increase. Finally, we can also see that Baby Boomers were the only generation to have a decrease in homeownership rates within this time period, this could be due to aging populations selling their home to move in with family or into care homes. In the future, we would expect to see the Baby Boomer home ownership rate continue to fall. Overall, we see that Millennials and Gen Z have the largest increase in home ownership rates within this time period.

Table 2: Home Ownership Rates (%) for ages 25-40 by Generation

Age	Baby Boomer	Gen X	Millenial	Gen Z
25	31.25654	26.84121	27.38984	28.95171
30	50.61442	48.34606	42.27259	NA
35	60.50193	58.64897	55.59943	NA
40	67.73174	63.52479	61.27793	NA

This table shows the homeownership rates at certain ages by generation. As the oldest members of Gen Z were only 26 in 2023, their is no data available in the later ages. We can see that Baby Boomers have the highest home ownership rate at all ages while Millennials have the lowest across all ages except 25. While Gen Z only has data at the age 25, it has the second highest home ownership rate at this age. There does appear to be a trend of homeownership rates dropping at all ages as the generations get younger; however, Gen Z could disrupt this trend if their rates continue to be high. Without further data on Gen Z, we cannot know if this trend will continue or Gen Z will outperform Millennials and/or Gen X.

Figure 1.
Home Ownership Rate By Year

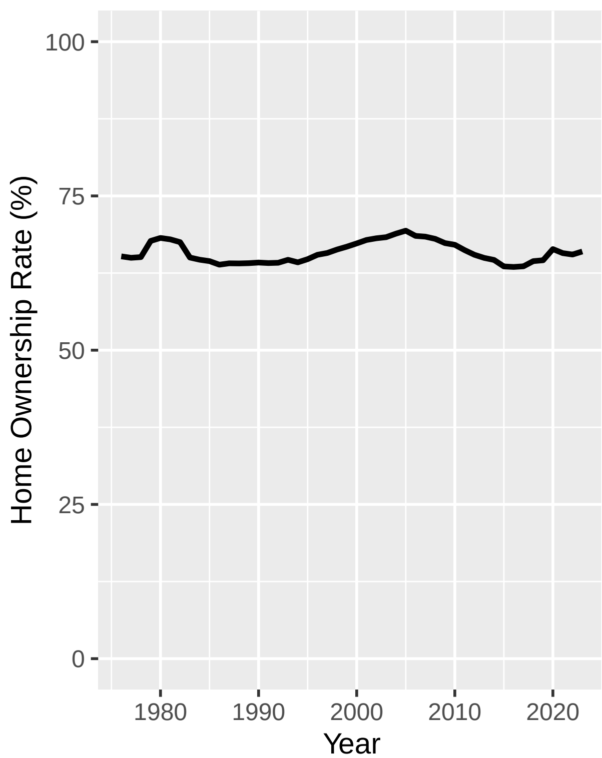
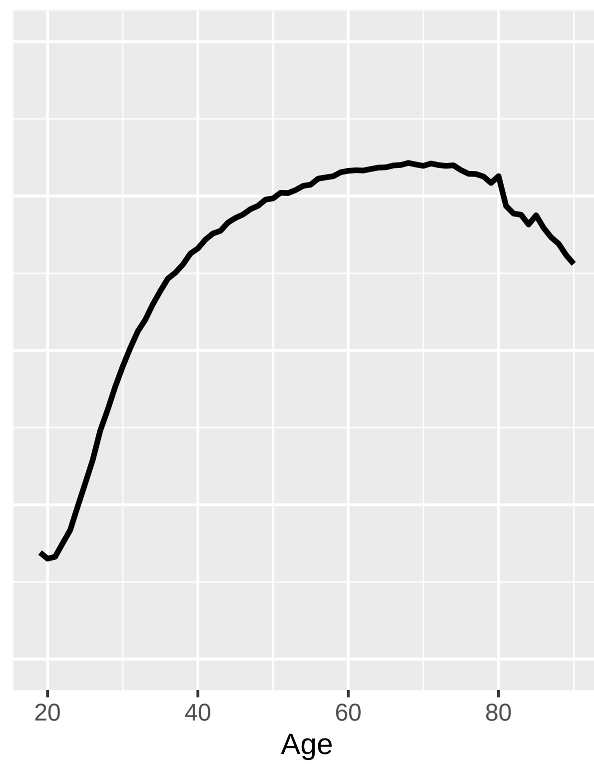
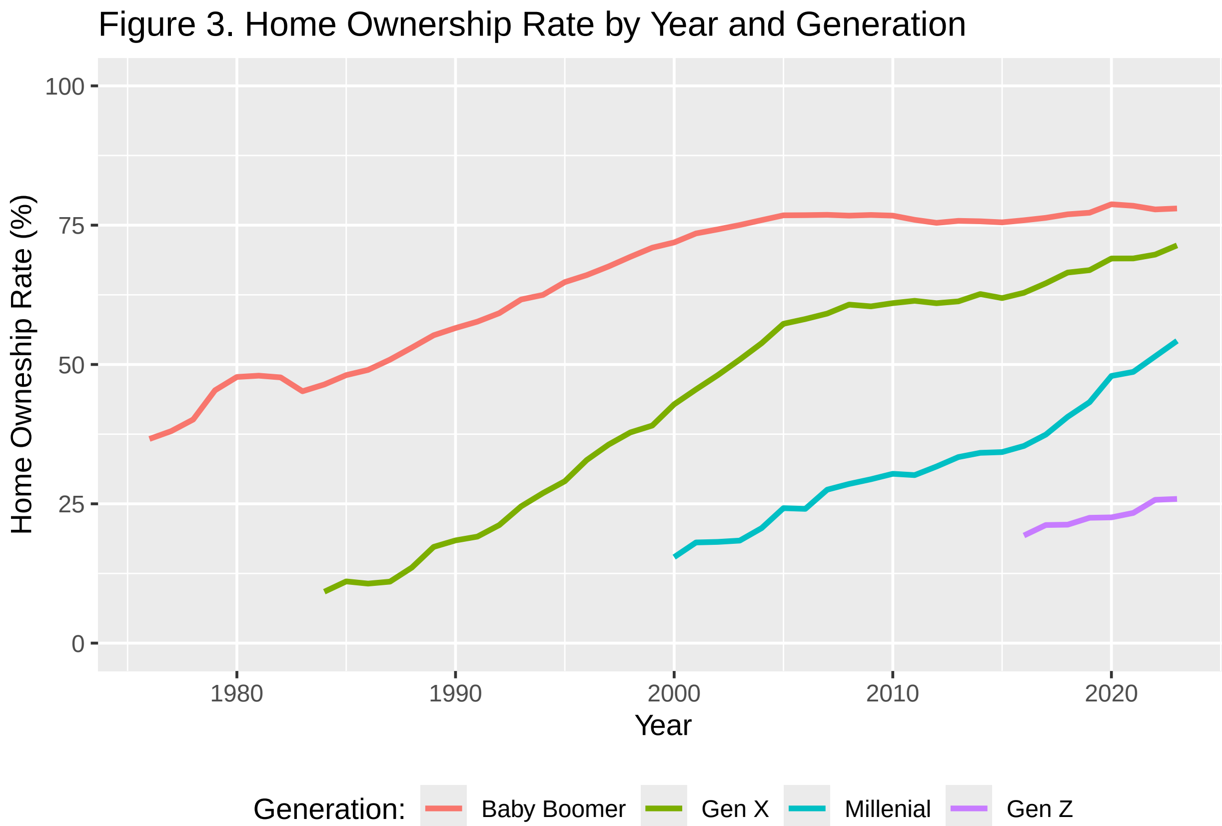


Figure 2.
Home Ownership Rate By Age

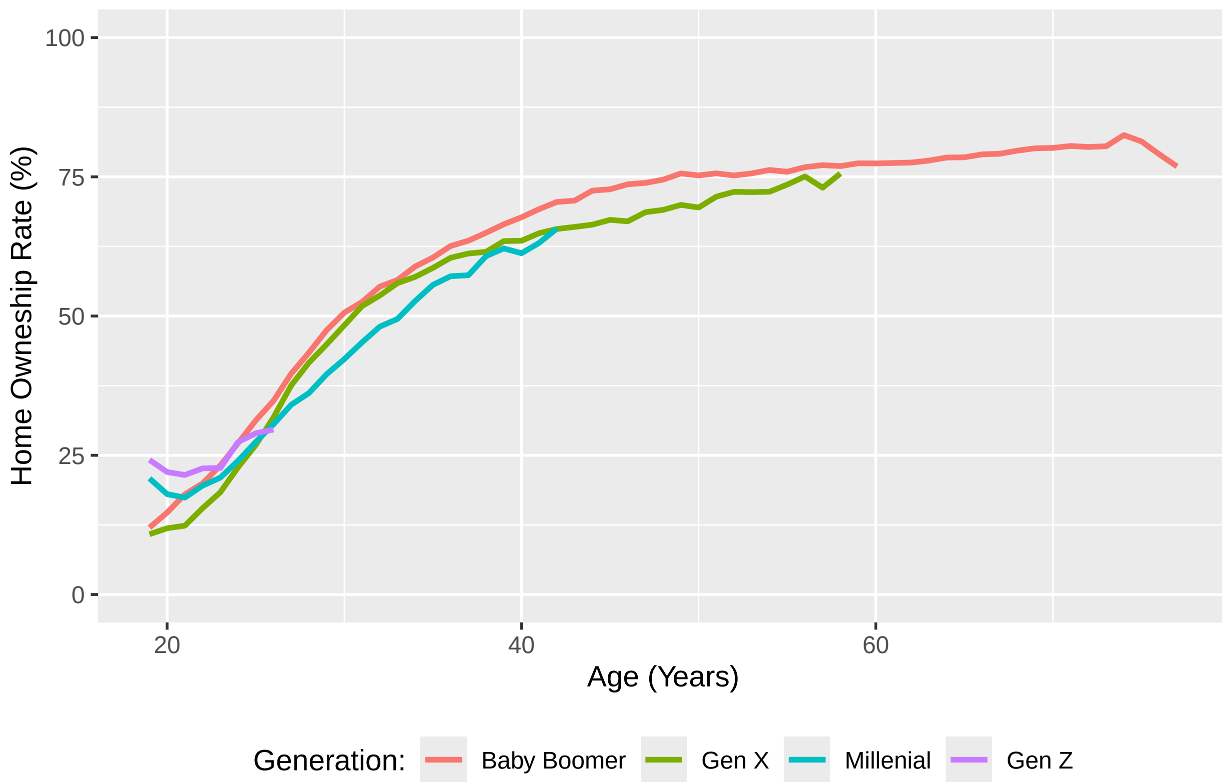


These plots show overall trends of home ownership by year and age. We can see that home ownership rates tend to stay fairly similar over time. This makes sense as outside of large economic shifts we would expect to have the proportion of people owning homes stay fairly constant as younger people buy houses and older people move out.We can also see that home ownership rates by age increases over time and then starts to flatten out around 50, before starting to decrease at around 80. Again, this would make sense as people would want to buy homes once they have a stable income and career which would likely occur in the age range 25-40. Then since most people who had the means to buy a house or wanted to buy a house likely already did by 50, the rate starts to flatten. Finally, once a person is older, they may start to sell their homes and move in with family or care home. Interestingly, we also see an outlier at around 19 where the rate starts off higher. The slightly elevated rate at 19 could be due to inheritance where an individual is not legally allowed to own a property until they reach age of majority.



This plot shows the trend of home ownership by year between each generation. We can see that Baby Boomer rate is has flattened significantly since around 2005 and the Gen X was starting to flatten around 2005 as well although it has started to increase again since around 2015; the sub prime mortgage crisis occurred between 2007 and 2010 and could have contributed to this decrease in the rate of change of the Gen X home ownership rate. For Millennials, we can see that the rate has been increasing and is starts to increase faster around 2015. Finally, for Gen Z there is not a lot of data, but we can see that it is starting to increase and in 2020 it had the rate started to increase faster.

Figure 4. Home Ownership Rate by Age and Generation



This plot shows the trend of home ownership by age between each generation. The plot provides a good comparison of home ownership rates by generation. We can see that Baby Boomers have consistently outperformed all other generations past the age of 26 and we can also see that Millenials have under preformed all other generations past the age of 26. The Gen X rate tracks fairly closely to the Baby Boomer rate before it starts to under preform in the 30s age range and onward. Gen Z out preforms both Millennial and Gen X for the age range where data is available, while the Gen Z line flattens towards the age of 26, without further data we do not know whether Gen Z will outperform or under preform other generations. Low intrest rates from 2019-2021, when older members of Gen Z would have started thinking about buying a home, could have contributed to the high home ownership rates of Gen Z relative to the other generations.

To see the interactive visualization of home ownership rates by state between generations that is analyzed in the following paragraph click here.

The map shows home ownership rates by state and generation. We can see that Baby Boomers have fairly high home ownership rates across all states. The states with the lowest home ownership rates for Baby Boomers are California, New York, and Hawaii. For Gen X, we see lower home ownership than Baby Boomers. We also see that Midwestern states and non-coastal states tend to have higher home ownership rates, which could be due to affordability. For Millennials, we can see again that home ownership is lower than Gen X, but we also see that the states with the lowest home ownership rates are similar to the states with the lowest home ownership rates for Gen Z. Finally, for Gen Z the states with the lowest home ownership rates are fairly similar to Millenials and Gen X, but surprisingly Florida is one of the states with the highest home ownership rate which is unique to Gen Z.

To see the interactive visualization of home ownership rates by county between generations that is analyzed in the following paragraph click here.

Similar to the previous map, this map shows home ownership rates by geography and generation. This map includes counties that were identified in the dataset, many counties were not identified in the dataset as their population was below a threshold were there are concerns about deanonymizing the data. All counties that were not identified in the dataset and any county with less than 50 observations in any generation are combined into the state and labeled “State w/o Counties”. Similar patterns can be seen to the previous state map; however, the difference between urban and rural home ownership rates are more obvious in this map, with rural areas tending to have higher home ownership rates across all generations.

Modelling

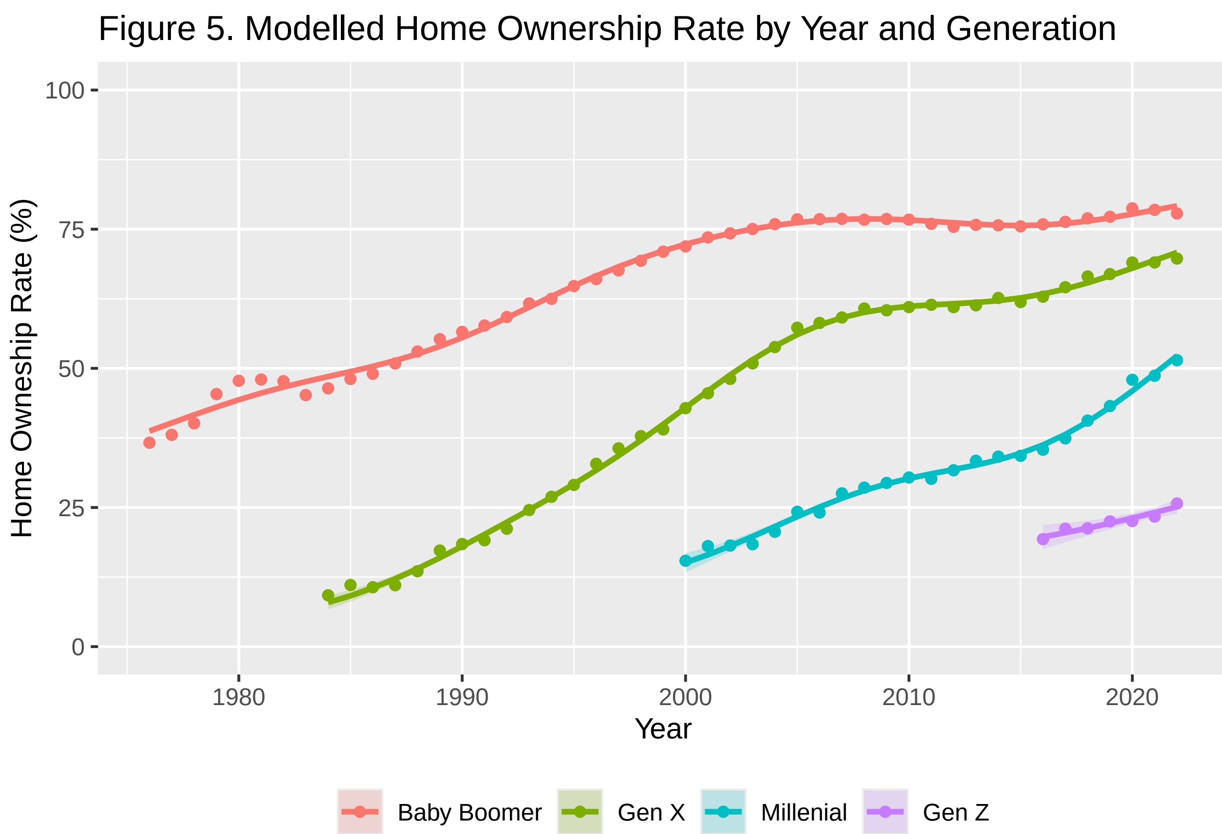
As described in the methods section, the goal of this section is to create a model that can predict the home ownership rates of a generation at a given year; the trend that we want to model can be seen in figure 3. The first GAM used year with an interaction for generation as predictors.

Table 3: GAM fixed terms

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.9167899	0.0027807	329.696233	0e+00
GENGen X	-1.0508553	0.0162941	-64.492814	0e+00
GENMillenial	-2.1617245	0.1608824	-13.436676	0e+00
GENGen Z	-2.5624865	0.5228998	-4.900531	1e-06

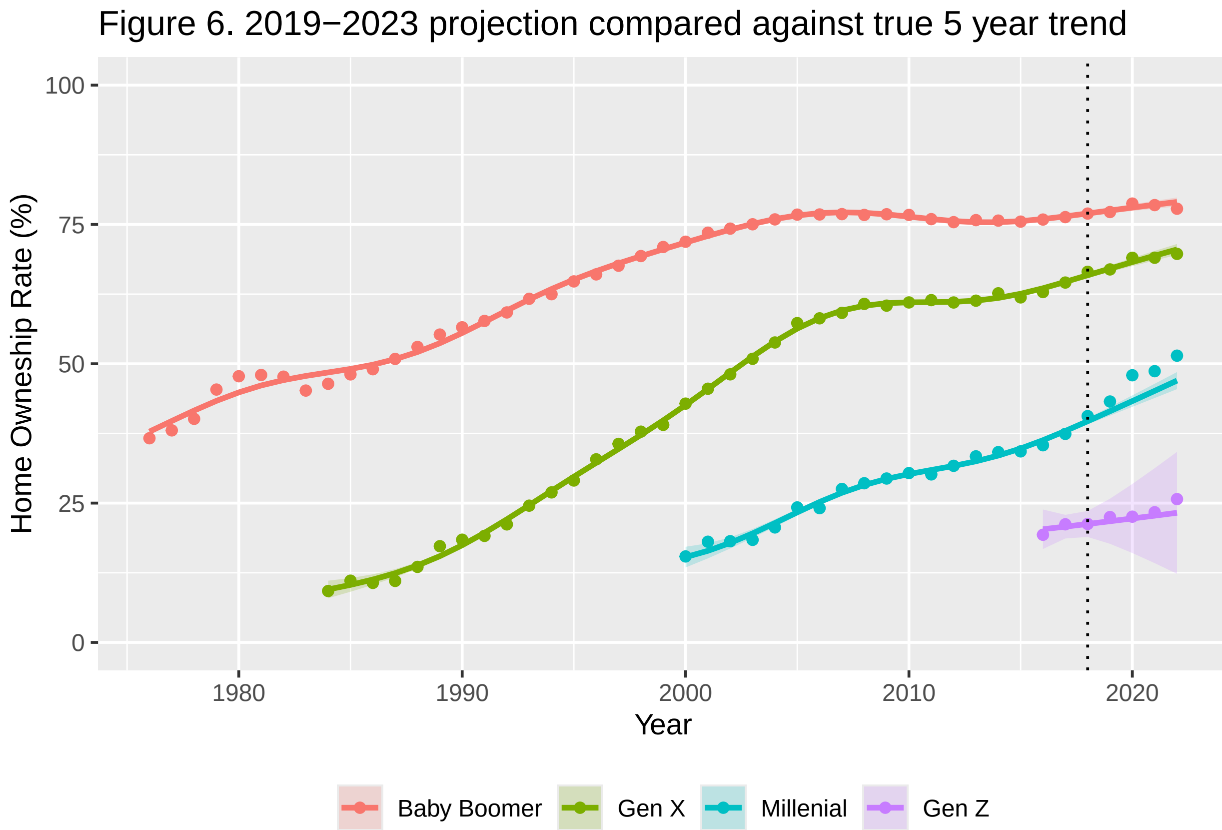
Table 4: GAM smooth terms

	edf	Ref.df	Chi.sq	p-value
s(YEAR,GEN)	17.66761	26	112578.2	0



This plot shows the fit of the GAM using year as a smooth term with generation as a smooth interaction. The observed ownership rates are shown as the points on the graph. The model achieves a very good fit to the data with 99.7% deviance explained and 0.996 adjusted R-squared. This is not unexpected as the model is producing almost an exact fit to the observed data; using this model beyond the years collected in the data set will likely produce unreliable predictions, especially as the year gets further from the observed years in the dataset. However, the model does show that both the generation intercepts and the smooth term where significant at $p<0.001$ showing that there is likely significant differences between the trends of home ownership between generations.

While the GAM will likely be unreliable far outside the 1976-2023 year range, we can test how well the model preforms on predicting home ownership trends five years into the future by withholding the last five years (2019-2023) in the dataset and training the model only on the remaining years.



This plot shows the GAM that was trained without the last five years (2019-2023) and then shows what the model predicts the trend to be in this year range. The true trend is shown as points. Within the year range 2019-2023, the model has a mean absolute error of 0.0143 (where the home ownership rate is on the scale 0-1). We can see on the plot that for Baby Boomers and Gen X, the mode ls predicted trend is extremely similar to the observed trend; however, for Gen Z and especially Millennial, the models predicted trend is not nearly as accurate. This is especially notable in the Millennial predicted trend where the observed home ownership rate for 2020-2023 are outside the 95% credible interval for predicted values. This implies that the model is not very accurate for predicting the five year trend for millennial home ownership, but the model does accurately predict the trend for Baby Boomers and Gen X.

While using only past data on home ownership rates between generation to predict future home ownership rates between generations would make a simple model that could be used to predict easily, it is unlikely any model that does not incorporate more information would be able to make accurate predictions farther into the future. Instead, a model that uses broad information that may relate to home ownership rates such as 30 year fixed mortgage rate, annual median home price, and average salary within a generation may be provide a model that is better at predicting home ownership further into the future.

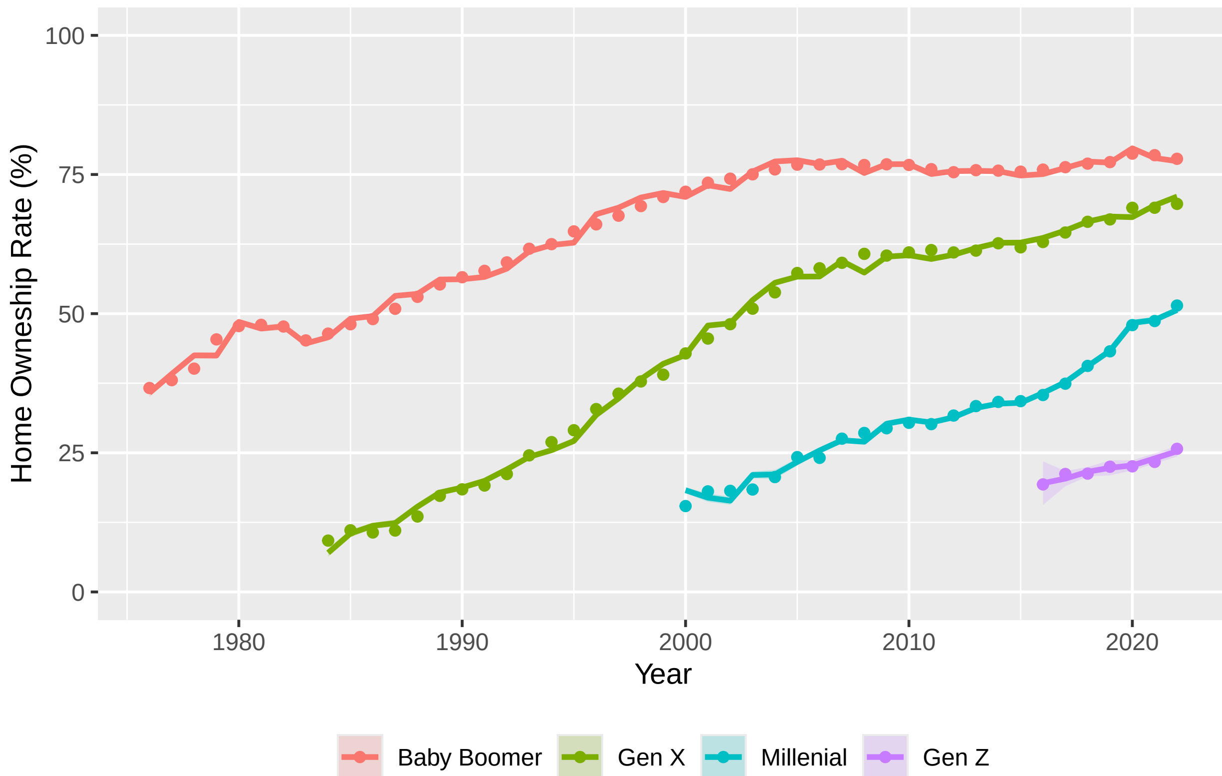
Table 5: GAM fixed terms

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.515237	0.0100637	51.19778	0
GENGen X	-0.735849	0.0118412	-62.14291	0
GENMillenial	-1.275487	0.0141501	-90.13983	0
GENGen Z	-2.025252	0.1367432	-14.81062	0

Table 6: GAM smooth terms

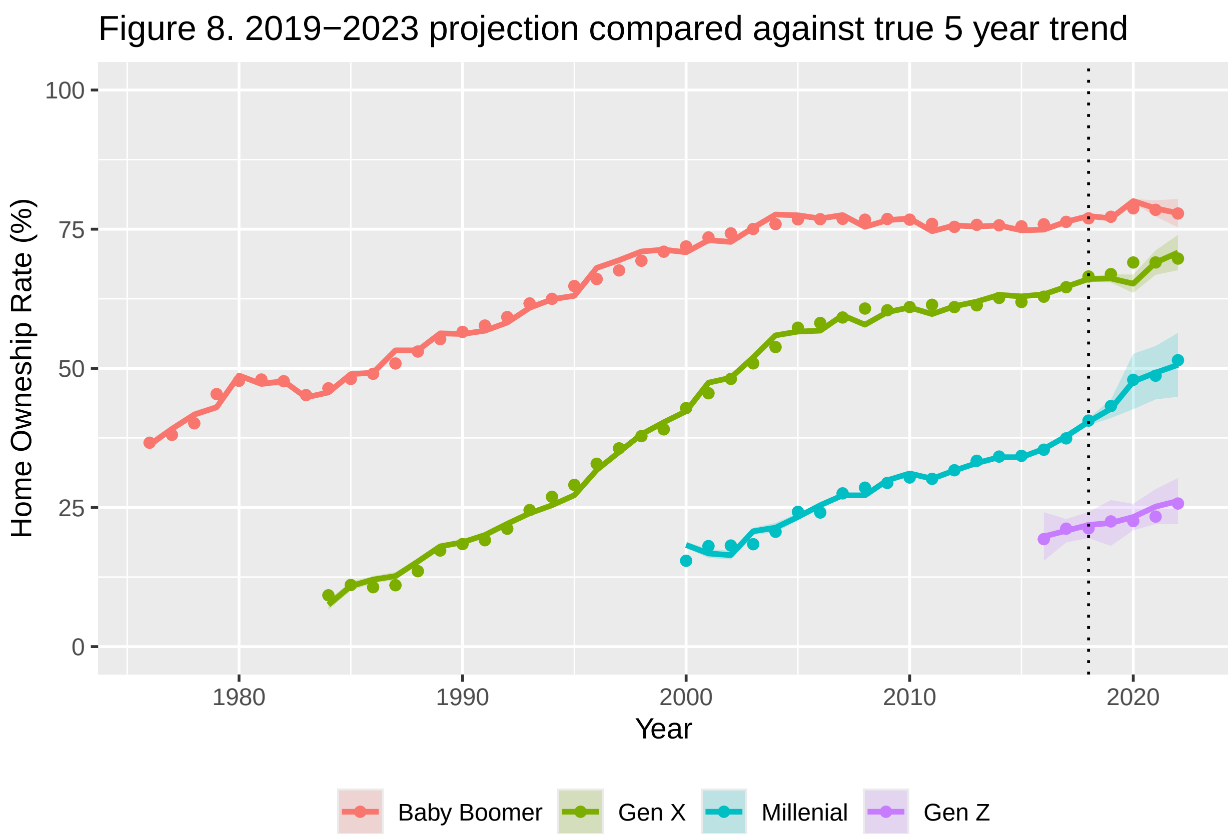
	edf	Ref.df	Chi.sq	p-value
s(INCOME,GEN)	9.734995	12.000000	12861.289	0
s(PRICE)	8.896411	8.995713	1270.601	0
s(INTREST)	8.734001	8.977514	3128.173	0

Figure 7. Modelled Home Owneship Rate by Year and Generation



This plot shows the fit of the GAM using average income as a smooth term with generation as a smooth interaction as well as median home price and 30-year fixed mortgage rate as smooth terms without interaction. To produce this plot, the average income for a generation, mortgage rate, and median home price were used as predictors to calculate home ownership rate for each year in 1976-2023. The observed ownership rates are shown as the points on the graph. The model achieves a very good fit to the data with 99.5% deviance explained and 0.993 adjusted R-squared. Unlike the previous models, this model does not directly fit the trend over time, rather it uses information about interest rates, home price, and income at a given time and predicts home ownership rate. Ideally, this would produce a model that would be capable of predicting future home ownership rates if data on future income, interest rates, and home prices was available. Like the previous model, the generation intercepts and all smooth terms where significant at $p<0.001$ showing that there is likely significant differences between the trends of home ownership between generations.

Similar to the previous GAM, we can test how well the model preforms on predicting home ownership trends five years into the future by withholding the last five years (2019-2023) in the dataset and training the model only on the remaining years.



This plot shows the GAM that was trained without the last five years (2019-2023) and then shows what the model predicts the trend to be in this year range. The true trend is shown as points. Within the year range 2019-2023, the model has a mean absolute error of 0.0081 (where the home ownership rate is on the scale 0-1). We can see on the plot that for all generations, the models predicted trend is quite accurate to the observed data. The only observed home ownership rate that falls outside the models 95% credible interval is the prediction for 2020 for Generation X; all other predictions are very close to the predictive trend or within the 95% credible interval of the predicted trend. We can see that this model is more reliable in predicting future trends than the simpler model, that only uses year and generation, by comparing the fit shown in the plot and the mean absolute error values. However, this model is much more difficult to use in aqquiring prediction for the future as to do so you would have to already have predictions for average income, mortgage rates, and median home prices; aqquiring data for future values of these variables is a challenging problem on its own. While this model may be more difficult to use in prediction, the accuracy of its future trend prediction shows that interest rate, household income, and home price, and generation are all predictive of home ownership rates.

Comparing between the two models, the first model was slightly better at fitting the true trend within the year ranges that were in the training data; however, for year ranges outside of the training data, the second model that used broader housing information gave more accurate predictions in the withheld five year range. While the first model could be used to easily predict future home ownership trends, the accuracy of these predictions, especially as the predicted years go far beyond the years used in training, will become unreliable. The second model could potentially be used to predict future home ownership trends, but this may be difficult as information on future mortgage rates, income within a generation, and home prices would be neccessary. Overall, both models show that home ownership trends between generations are different, and that broad housing information can be used to gain fairly accurate predictions of home ownership rates; particularly, average income within a generation, mortgage rates, and median home price are all predictive of the home ownership rate for a generation in a year.

Conclusion

The analysis compared home ownership rates across generations by investigating trends in home ownership rate over time, age, and geography.

The trend of home ownership rate over years between generations was similar, with fast growth followed by a flattening of the curve; however, some differences between home ownership rate over year trend could be seen, such as the increase in Gen X home ownership rate slowing down between the years 2005 and 2015, potentially due to housing market conditions in this period. While some differences were present, the overall trends appeared similar, showing that the trend home ownership rate over time has not significantly changed between generations.

The trend of home ownership rate over age between generations was also similar. Although some deviations could be seen in the trends between generations. Millennials tended to have the lowest home ownership rates at all ages when compared to the other generations, while Baby Boomers tended to have the highest rates at all ages. Gen X tracked closely with Baby Boomer between the ages 19-35 but started to under preform at later ages. Surprisingly, Gen Z was had the highest home ownership rates in ages 19-23 and the second highest in ages 23-26 with Baby Boomers as the highest in that age range. The trends between each generation being similar and the high home ownership rates of Gen Z indicate that home ownership has not increased or declined significantly between generations despite Baby Boomers and Gen X tending to have higher rates.

The analysis of home ownership rates by state and by county showed that home ownership rates tended to be similar across geography between the generations. Home ownership rates in less populous states and counties, such as Midwestern states or non-identified counties, tended to be higher across all generations; although there were some interesting differences, such as Florida being one of the states with the highest home ownership rate for Gen Z compared to all other generations were Florida was not one of the states with high home ownership rate.

The modelling of home ownership rates showed that past home ownership rate trends over time could be used to predict future home ownership rates within a five year time period, although these predictions were not as reliable for the Millennial and Gen Z generations. Additionally, the modelling showed that average income within a generation, 30-year fixed mortgage rates, and median home price were all predictive of home ownership rates within a generation, and using them could provide very accurate predictions of future home ownership trends if data on these variable was available for the years that we want to predict. Finally, the modelling showed that there was significant differences in the trends between generations confirming the initial analysis of the plot that showed small differences in the trends such as the Gen X rate slowing down between 2005 and 2015.

Overall, the analysis showed the the general trends of home ownership rates by year, age, and geography have remained similar between generations, although there are significant differences in specific trends over time that can be seen in the initial analysis and modelling. While home ownership rates do not appear to have declined significantly between generations, there is a pattern of home ownership trends flattening as the generations get younger. This analysis focuses on home ownership rates which do not necessarily reflect general housing market conditions, so it cannot give an accurate picture of the housing market or housing affordability. While some additional information such as mortgage rates and home prices was incorporated into the modelling, further analysis that included more economic data would provide more cohesive information about housing market between generations.

Citations:

IPUMS: Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry. IPUMS CPS: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2023. https://doi.org/10.18128/D030.V11.0

Generation Ranges:

[1] Dimock, M. (2019, January 17). Defining generations: Where millennials end and generation Z begins. Pew Research Center. https://www.pewresearch.org/short-reads/2019/01/17/where-millennials-end-and-generation-z-begins/

Home Ownership Definition:

[2] Definitions and Explanations. United States Census. https://www.census.gov/housing/hvs/definitions.pdf

Home Ownership as a voting issue:

[3] O'Donnell, K. (2022, August 2). Young people are pissed off: Housing crush sours millennial voters. Politico. https://www.politico.com/news/2022/08/02/housing-millennials-biden-economy-00047704

Historical Intrest Rates:

[4] Freddie Mac, 30-Year Fixed Rate Mortgage Average in the United States [MORTGAGE30US], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MORTGAGE30US, April 26, 2024.

Historical Median Home Price:

[5] https://dqydj.com/historical-home-prices/

Geojson for states:

https://rstudio.github.io/leaflet/json/us-states.geojson

Geojson for counties:

https://gist.github.com/sdwfrost/d1c73f91dd9d175998ed166eb216994a