COMP4447 – Data Tools 1 Final Project

Authors: Sammantha Firestone and Jacquelyn Noyes

**Datasets and Motivation**

Currently in the United States, one of the largest economic barriers that exists is the unaffordability of housing. Historically, the average cost of a single-family home was around five times more than that of the average yearly household income; however, in recent years, that ratio has increased to the average house costing over seven times more than the average household income.

The motivation of this analysis was to further understand the disproportionate increase of housing prices compared to household incomes. To accomplish this, we pulled from multiple data sources to construct a set that fit our needs.

The data for median household income was collected from the [US Census](https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-income-households/h08.xlsx). The data is organized based on median income per year, from 1984 to 2021, for each state in the United States. There is also a row that shows the median income throughout the years for the entire country.

The data for the average price of a house was collected from Zillow. One dataset listed the [median price](https://files.zillowstatic.com/research/public_csvs/zhvi/Metro_zhvi_uc_sfrcondo_tier_0.33_0.67_month.csv?t=1664126527) of a house based on year and region of the United States. The other dataset contained an algorithm for [predicting price](https://files.zillowstatic.com/research/public_csvs/zhvf_growth/Metro_zhvf_growth_uc_sfrcondo_tier_0.33_0.67_month.csv?t=1664126527) increases and/or decreases based on previous trends.

For this analysis, we will clean and combine the data taken from the above datasets to create a singular data frame that includes median household income and average house price based on year and state. We will then work to visualize the growing gap between household wages and housing market prices.

**Literature review**

Unsurprisingly, there have been numerous other investigations of this topic.

<https://www.jchs.harvard.edu/blog/price-to-income-ratios-are-nearing-historic-highs>

<https://anytimeestimate.com/research/housing-prices-vs-inflation/>

The above sources have looked at other factors affecting inability to afford a house (tuition, consumer price index increase) rather than just the relationship between household income and housing prices.

**Results**

Based on the data collected, cleaned and analyzed, from the Zillow ZHVI typical housing prices and the US Census median household income datasets, it is clear that the typical price of homes are unproportional to what the average American's household makes. Many banks and financial websites state, that individuals and families looking to purchase a home, should follow a 28% | 36% "rule of thumb". This "rule of thumb" states you should spend 28% or less of your monthly gross income on your mortgage payments, and no more than 36% of a household income should go towards **all** debt, *including* mortgage payments.  
From our Median Household Income to Typical House Price Ratio by Region box-and-whisker plot, we already know that not all Americans can follow this rule, even when we use the capped-off 36% ratio (this is assuming that these families have no other debt!).  
  
A housing bubble occurs when the price of homes rises at a rapid pace. Many things can cause a housing bubble like increase in demand, limited supply, and emotional spending. These bubbles are often fueled by prolonged periods of sub-normal interest rates and the race to purchase property. This phenomenon is called a "bubble" because at some point, it bursts. This is what happened in the 2008 Housing Market crash. The burst began in 2007 when interest rates began to climb, and property demand fell and then continued threw 2010.  
We chose to look at the housing debt-to-income ratio in 2006 because this was the peak of the housing bubble before it bursted. Comparing it to 2021, we can feel slightly better about where we are financially as a country because it appears Americans did not take out as much property debt as they did during the 2006 housing bubble. However, we should not sit too comfortably. There are many other factors that goes into a bubble burst / housing market crash that are not reflected on here, like interests rates, student debt, and the global economy as a whole.  
  
Some economists believe this bubble will not burst, and property values will continue to increase over the years without a sudden decline, and instead it will simply "correct" itself. Other economists do not believe this will be the case, as we finish 2022 interest rates have increased 2% - 3% since the beginning of the year, and houses are less likely to be sold more than their real worth.  
Overall, we will not truly know whether we are in a housing bubble until it bursts and the repercussions pass us. Regardless, the current typical price of a home for the average American is disproportionate to their income. This is an serious issue as many Americans could face financial disaster if or when a recession hits and the price of essential goods significantly rise, leaving them with tough financial choices.

**Suggested next steps**

As highlight throughout this notebook, there are many factors that affect the housing market. Next steps that can be performed to get a better understanding of the current market goes as follows:

1. Integrating interest rates into the housing prices dataset
2. Research and understand the typical debt the average American faces (i.e. student loans)
3. Integrate these debts into the debt-to-income ratio touched upon this notebook
4. Improve this notebook by taking housing price sold data rather than Zillow ZHVI (because this takes the current economy into consideration in the typical housing prices)

Other factors may be considered in this analysis that we did not touch upon. Real estate economics is a multi-faceted topic that is ever-growing and ever-changing. There will never be a perfect way to model and analyze it, but with these basic steps, we can gain an understanding of where are now and where we have been.