Real Estate Valuations

Data Science Project

Clare Kenney
Marquette University
Milwaukee, Wisconsin
clare.kenney@marquette.edu

ABSTRACT

Real estate and the housing markets have been a topic of conversation for quite some time now. Historically speaking, the real estate market has been a major determinate of our global economy. In 2008, we saw this firsthand with the crash of the stock market. Arguably, the main cause of this crash was the banks giving housing loans to people who could not afford their mortgages and overtime, this lead to major banks declaring bankruptcy as too many people defaulted on their loan payments. Economist and real estate agencies have dedicated their time to studying the housing market and the trends in real estate. The varying values of different properties is determined by three main forces: social ideals, the economy, and governmental regulations. Through research, we can see that there are many factors that are determinates of real estate valuations including the age distribution of the area, educational opportunities and desire for privacy, but does the distance to mass transportation or the age of the house play a role in the price valuation? Do we see a correlation between the number of convenience stores in the area and whether someone buys a house in that area? Using a data set that collects the market historical of real estate valuations, I created graphs and charts to help determine if the house price of unit area was correlated with any of the data in the chart including, house age, distance to mass transport, and number of convenience stores in the area. From my visualizations, my data supported the idea that both age and distance to mass transportation negatively correlated with housing valuations. As the age of the house or the distance to mass transportation increased, the price valuation of the house decreased. My visualizations also supported the idea that number of convenience stores does not have an impact on whether or not someone buys a house in the area.

ACM Reference Format:
Clare Kenney. 2018. Real Estate Valuations: Data Science Project. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

The real estate market is a complex industry that can be affected by many factors. It is a hard market to predict as it is always changing and certain city's housing markets are different from others. Most recently, we can see the varying fluctuations in the health of the housing market as Denver's market has specifically has slowed down. In a *New York Times* article, realtor Steve Danyliw weighed in on the Denver housing market. He claims, "The local economy is still fantastic, all the fundamentals are there, but obviously, wages are not keeping pace. As the market continues to move up, buyers are being pushed out." Similar slowdowns are being experienced in New York, Seattle, and San Francisco—prices have been increasing at a rate faster than income gains. [1] These fluctuations in the housing market can have a direct impact on the economy. Just as the housing market fluctuates, so does the various trends and valuations that determine real estate prices and costs.

The trends and valuations of real estate have changed over time as our society has developed and new standards of living have become more ideal than others. The banks play a role in real estate valuations, however, it is important to note that it is the buyers that give value and significance to real estate. In *The Functions and Purposes of Appraisals* article, it says, "Real property has significance only as it fulfills man's need and desires. Man's utilization of real estate gives it character; man's collective desire for real estate gives it value."[2] Without the desire or demand for real estate, the industry would have no value or pricing power—houses close to a park could not price higher than houses further from a park if there was no desire to be near a park or apartments with a view wouldn't be able to be more expensive than apartments without a view if the scenery wasn't desirable.

Through data wrangling, data cleaning and regression analysis, I was able to create graphs and charts that help visualize if the age of houses and its proximity to mass rapid transportation have an impact on real estate prices. I was also able to visualize if the number of convenience stores in the area had an impact on if someone bought a house or not. I found, through my visualizations, the following findings:

- As the age of the house goes up, the price of the house decreases.
- As the distance to mass rapid transportation increases, the price of the house decreases.
- The number of convenience stores in an area has no impact on whether or not someone buys a home.

. Through the articles I have read about the real estate industry and the varying real estate valuations, I have found that banks and the real estate industry are dependent on one another and the economy and the housing market are closely connected—specifically when the housing market is suffering, the economy will too. I have also

Conference'17, July 2017, Washington, DC, USA

[1]Casselman, B. 2018. Housing Market Slows, as Rising Prices Outpace Wages. The New York Times
[2]Education Committee. 1951. The Appraisal of Real Estate. American Institute of Real Estate Appraisers. Chicago, IL

found that there is not just one or two factors that play into the pricing of real estate, but rather multiple factors that contribute to real estate costs and valuations. These findings affect everyone around the globe and the trends and standards ultimately show that it is the social ideals that influence the real estate costs.

2 LITERATURE REVIEW

I read scholarly articles about real estate and real estate valuations to help further my understanding on the topic and formulate my research questions. The literature I read not only talks about real estate valuations and factors, but it also discusses the housing market as a whole and the impact it has on the economy.

2.1 What is Real Estate?

A key to understanding the topic of real estate is to understand what real estate means and the terminology that is often associated with it. According to Webster's dictionary, real estate is "property in buildings and land." When you purchase a home, you are buying real estate. When you rent a building space, you are renting a piece of real estate. A couple key terms that are often associated with real estate include: appraisal and mortgages. An appraisal can be defined as "an estimate of value. It is usually a written statement of market value, or value for loan purposes, or value as defined by the appraiser, of an adequately described parcel of property as of a specified date."[3] Another name for appraisal is valuation or property valuation, which can be influenced by multiple factors that are discussed below. Another term that is important to understand is mortgages. A mortgage is an agreement between the bank and yourself that allows you to borrow money from the bank to buy a house or land and in exchange, you have to make monthly payments to the bank until the loan is paid off. Understanding what a mortgage is and the importance of it, is crucial in buying any sort of real estate as well as understanding the market. Real estate plays an important role in the economy as well in people's personal lives.

2.2 Banks and Real Estate

The banks and the real estate industry have a complex relationship. They both play major roles in our nation's economy and they are both reliant on one another to operate. When the real estate market is doing well, the bank benefit and when the real estate market is not doing well, the banks suffer. In the article *Banks and Real Estate Prices*, it says, "Banks make above average profits in the upswing phase of the real estate cycle but suffer high losses when the market turns."[4] These industries' close ties can be both a positive and a negative as we have seen through stock market highs and lows. The role the banks play in real estate is lending money to people who want to buy real estate in the form of a loan, in which the person borrowing money, pays the mortgage to the bank every month. This mortgage payment allows the person to continue living in the house as long as they pay the bank. The real estate industry and the banks are fully reliant on each other. In *Banks and Real Estate Prices*, it says, "The willingness of banks to provide funding for real estate purchases depends on the creditworthiness of their borrowers. Besides other factors, the creditworthiness of the borrowers depends on the development of real estate prices. Real estate prices, in turn, depend on the demand for houses which is influenced by the willingness of the banks to provide funding for real estate purchases." [5] This reliant relationship can become an issue as we have seen throughout history with housing bubbles and market crashes.

2.3 Housing Bubble

The relationship between banks and the housing market is an important aspect of our economy. However, the dependent relationship can also cause issues in the economy and housing bubbles can form. A housing bubble is defined as "a period in the real estate industry when house prices grow to above-average...until they can no longer be supported." This can be explained by simple supply and demand concept. If there are too many houses on the market, but not enough demand, the prices go down.[6] According to *Bubbles in Real Estate Market*, "Real estate bubbles may occur without banking crises. And, banking crises may occur without real estate bubbles. But the two phenomena are correlated in a remarkable number of instances ranging over a wide variety of institutional arrangements, in both advanced industrial nations and emerging economies."[7] Most recently, the U.S. economy had a housing bubble in the early 2000s that ultimately collapsed in 2008. The prices of houses reached a high point in 2006 and slowly began to decline. Other factors did play into the role of the housing bubble burst in 2008, but a large part of it was the role the banks played in the real estate industry.[8] The housing bubble is important to understand when discussing the most recent stock market crash, which was in 2008.

2.4 2008 Stock Market Crash

The most recent stock market crash in the U.S. economy was in 2008. A stock market crash is defined as "a rapid and often unanticipated drop in stock prices." A stock market crash can be caused by many events including economic crisis, catastrophic events or collapse of a long-term bubble. This major crises in the U.S. economy can be traced back to many places, but one of them being the increase in sub-prime mortgages. The housing prices grew exponentially and continued to grow until around 2007. In the article *Stock Market Crash of 2008*, it details the time line of events leading up to it. It says, "The Dow opened the year (2007) at 12,459.54. It rose fairly steadily throughout most of the year despite concerns about a slowdown in the over-heated housing market. In fact, there had been warning signals as early as 2006 that the housing market was starting to falter. The Commerce Department warned on November 17, 2006, that October's new home permits were a whopping 28 percent below the October 2005 rate. Housing prices were falling in 2006, triggering the default of sub-prime mortgages. But government officials didn't think the housing slowdown would affect the rest of

[3] Education Committee. 1951. The Appraisal of Real Estate. American Institute of Real Estate Appraisers. Chicago, IL

[4] Holt, C. Banks and Real Estate Prices. Swiss National Bank

[5] Holt, C. Banks and Real Estate Prices. Swiss National Bank

[6] Smith, L. 2018. What Is a Housing Bubble?. SmartAsset, https://smartasset.com/mortgage/what-is-a-housing-bubble

[7] Herring, H and Wachter, S. 2005. Asset Price Bubble: The Implications for Monetary, Regulatory, and International Policies. MIT Press, Cambridge, Massachusetts

[8] Amadeo.K. 2016. Stock Market Crash of 2008.https://02038.com/wp-content/uploads/2017/03/Stock-Market-Crash-of-2008.pdf

the economy."[9] Government officials were wrong as we can see in 2008. Major banks like Lehman Brothers filed for bankruptcy and over the course of September the stock market fell, specifically on September 29, 2008, the stock market "fell 777.68 points," which was the "largest point drop in any single day in history."[10] It is hard to pinpoint an exact date on which the stock market fell, but according to the article *Stock Market Crash of 2008*, it was on September 29, 2008 when the stock market prices dropped drastically.

The relationship between the banks and real estate industry is an important aspect in understanding the U.S. economy as it is estimated that real estate represents "two-thirds of this country's total wealth." The economy and the real estate industry are closely related to one another. This one article says, "One of the fundamental indications of the economic significance of real estate is the inseparable relation between the economic health of real estate and the economic well-being of the entire national economy."[11] Real estate prices and valuations can be due to the health of the economy, but there are also many other factors as well including social, economic, and political factors.

2.5 Determinants of Real Estate Values

Real estate valuation factors are factors that positively or negatively impact a products price of value. According to *The Appraisal of Real Estate*, "The value of real estate is created, maintained, modified, and destroyed by the interplay of the three great forces that motivate and characterize the activities of mankind." These three forces are "social ideals and standards, economic adjustments, and political or government regulations." These factors all play a role in the real estate market. Generally speaking, "it may be said that value is created in land and buildings by interrelated forces of social standards, economic adjustments and governmental regulations."[12] There is also a personal aspect when choosing a home that may drive someone to spend extra money to have a specific amenity or perk to it.

2.5.1 Social Factors. The social factors that contribute to determining real estate costs and values are based on society's ideals and standards. Some of these factors include: geographical age distribution, race distribution, desire for educational opportunities, privacy, size of family, and religious distribution. These are just a few factors that impact real estate prices. In fact, the area around the house can be equally, if not more, important than the house itself in determining real estate value. There have been multiple studies done that show that certain features of an area affect the real estate valuations in the area significantly.

There were four studies done in four of the articles I read about the impact of churches, parks, schools and recreational amenities in an area on house prices. To determine "the impact of neighborhood churches on residential property values," a study was conducted on 5,000 residential property transactions in Henderson, Nevada. The evidence of the study suggested that "neighborhood churches are amenities that enhance the value of neighborhood residential property." In other words, the presence of a church can increase residential property value.[13] Another study about the impact of a park on the surrounding area's property values determined that the "analysis of property sales in the vicinity of 1,294-acre Pennypack Park in Philadelphia indicates a statistically significant rise in land value with closeness to park."[14] Another study was conducted to determine the impact of the quality of schools in the area on house prices and it found "that public school quality has a very large impact on real constant-quality house prices."[15] Lastly, a study conducted to determine if the proximity to recreational amenities had any effect on real estate value determined that housing values increased in an area close to a shopping center or other recreational area.[16] Another factor that can influence real estate prices is transportation in an area.[17] This idea led me to formulate one of my research questions stated below.

All of these factors are based off of society's ideals and standards however, depending on the stage of life you are at and your values, there may be specific characteristics or qualities you look for in your home or apartment. For example, if you are young in your 20s with no kids, you may care more about the age distribution of the area you live in rather than the education opportunities the area has. If you are married with kids, you may look for a place that is close to a park and where other families live rather than a place where the average age is 24 years old. Our own values ultimately determine our housing choices, but it is society that can determine if something is valuable.

Societal ideals are not only about the area in which a house is in, but also about the house itself. In the article, *Determinants of Real Estate Values*, it claims that the "services provided by any particular house clearly depend upon a great variety of physical characteristics of the house and its grounds. Possibly the most important of these characteristics of the house and its ground."[18] The house itself goes into the valuation of the property. If the house is recently renovated or is built on a large lot, it may go for more money than one that has minimal land and is a smaller house. Other features that are used to determine house values are how big the house is, parking, heating, electrical, how many stories it is, presence or absence of family area, and the house's general condition. This idea of social factors correlating to house prices leads me to my research questions:

- Does the age of a home or the distance to mass rapid transportation affect the price of the home?
- Does the number of convenience stores in the area affect the amount of people that will purchase a home?

[9] Amadeo.K. 2016. Stock Market Crash of 2008.https://02038.com/wp-content/uploads/2017/03/Stock-Market-Crash-of-2008.pdf
[10] Amadeo.K. 2016. Stock Market Crash of 2008.https://02038.com/wp-content/uploads/2017/03/Stock-Market-Crash-of-2008.pdf
[11] Education Committee. 1951. The Appraisal of Real Estate. American Institute of Real Estate Appraisers. Chicago, IL
[12] Education Committee. 1951. The Appraisal of Real Estate. American Institute of Real Estate Appraisers. Chicago, IL
[13] Carroll, T.M., Clauretie, T.M., and Jensen, J. Living next to godliness: Residential property values and churches. The Journal of Real Estate Finance and Economics, Vol 12, 319-330
[14] Hammer, T.R., Coughlin, R.E. and Horn, E.T. 1974. The Effect of a Large Urban Park on Real Estate Value. Journal of the American Institute of Planners. Vol 40. p. 274-277
[15] Haurin, D. and Brasington, D. 1996. School Quality and Real House Prices: Inter-and-Intrametropolitan Effects. Journal of Housing Economics. Vol. 5. pg. 351-368
[16] Hammer, T.R., Coughlin, R.E. and Horn, E.T. 1974. The Effect of a Large Urban Park on Real Estate Value. Journal of the American Institute of Planners. Vol 40. p. 274-277
[17] Haider, M and Miller, E. 2018. "Effects on Transportation Infrastructure and Location on Residential Real Estate Values" *The National Academies of Sciences Engineering Math*
[18] Grether, David M. and Mieszkowski, Peter (1974) Determinants of real estate values. Journal of Urban Economics, 1 (2). pp. 127-145. https://authors.library.caltech.edu/83731/

.

2.5.2 Economic Factors. Social factors aren't the only factors that can affect real estate prices and values. Economic factors also play a major role in the price of a home and whether or not someone wants to purchase a home or not. Some economic factors that affect real estate's value include: wage levels, employment trends, tax loads, price levels, quantity, quality and location of natural resources, unionization of labor and interest rates. The economic factors have a large impact on where someone chooses to live because their personal income or situation can determine if they are able to live in that area or not. Someone may find an affordable house in a very expensive area with high taxes and interest rates, so these economic factors can sway people from purchasing homes. Some people may want to live in an area where the wage levels match theirs or a large majority of the people are employed. The difference between social factors and economic factors is that the economic factors can be a larger determinant on where someone can realistically live rather than where they want to live. [19]

2.5.3 Political Factors. Social and economic are large factors in real estate valuations, but political factors also play a major role. Some of the political factors that affect real estate's value include: "building codes, zoning laws, public health measures, social security measures, fire regulations, government housing, rent control, government guaranteed loans, national defense measures, priorities, allocations, credit controls, monetary and fiscal policies." [20] Political factors are determined by a community, state or even national government. These factors not only have an impact on the real estate's value, but they also have an impact on whether or not someone buys a home in the area. If someone is looking to build a large, modern house and the building codes in the area restrict that person from building their dream house, then a person may look in another area with looser building codes. Political factors, similar to economic factors, also have a major impact on whether or not someone can financially live in an area or not. If someone needs government housing, it will most likely be in a specific area and not many other options to choose from.

3 METHODS

I used a dataset found on UCI. This dataset gave me the different real estate transactions from the year 2012 to 2013. The data included the transaction date, age of the house, the distance from mass rapid transportation, the number of convenience stores in the area, latitude and longitude degrees and house price of unit area. This data needed to be wrangled to display only the columns needed for visualizations, so I condensed the table to only be the age of the house, distance to mass rapid transportation, number of convenience stores in the area and house price of unit area. I needed these variables, so I could compare the distance to mass rapid transportation, number of convenience stores and house age to the house price of unit area to determine if any of the variables had any impact on the house's price.

First, I looked at the distance to mass rapid transportation and house price of unit area to see if there was any correlation. I used a scatterplot to determine if there was any correlation because both of the variables were continuous variables. I calculated the line of best fit to get a proper understanding of the correlation between the two variables. Then, I used a bar graph to calculate the count of the amount of transactions for the number of convenience stores in the area. The bar graph helped visualize if there was any correlation between the number of convenience stores and whether or not someone purchased a home. I used another scatterplot to plot the house age and house price of unit since they are both continuous variables. I used the line of best fit to help interpret the data. Finally, I used the calculated statistical model summary table to show important values including p-value and the coefficients to help in my regression analysis. An example of the statistical model table I used for the Number of Convenience stores is pictured in Figure 3. You can see my methods and visualizations at https://github.com/clarekenney/DataScienceProject2. This is where you will find my csv files and link to jupyter notebook as well.

3.1 Analysis

My calculation of the statistical model table for the different variables in my project helped me make conclusions about the dataset. I first looked at the varying r-square values to make sure the r-squared was not 0, as that indicates error. However, it is important to note that I want to determine whether or not these variables are statistical significant and how the changes in the variables result in a change of the response (house price), so the r-value is not as important as the p-value.

The next value I looked at was the p-value because that determines whether or not to reject the null hypothesis. The null hypothesis is that the coefficient is equal to 0, which means there is no effect on the data from the variable. In other words, the null hypothesis states that house age, distance to mass rapid transportation and number of convenience stores have no impact on the price of the house. In all four of my statistical model tables of each variable, the p-value was equal to zero, which means that changes in the variable's value are related to change in the response variable, which the response variable in my research is the house price. To put it simply, the variable has an effect on the data that is given. In the figure below, the p-values for the different variables are shown.

	P>|t|
House Age	0.000
Number of convenience stores	0.000
Distance to Mass Rapid Transportation	0.000

Figure 1: Screenshot from the statistical model table of the p-values for the given variables

[19] Education Committee. 1951. The Appraisal of Real Estate. American Institute of Real Estate Appraisers. Chicago, IL

[20] Education Committee. 1951. The Appraisal of Real Estate. American Institute of Real Estate Appraisers. Chicago, IL

4 RESULTS

Through my research and dad visualizations, I have found that my data set supports the idea that the age of the house and the distance to mass rapid transportation does correlate with the price of a house. Generally speaking, as the age of the house increases, the price decreases and as the distance to mass rapid transportation increases, the price decreases. The correlation between the age of the house and the house price of unit area was not as strong as the distance to mass rapid transportation, but we can still see a slight, negative correlation. Figure 2 below shows the negative correlation of house age against the price of the house.

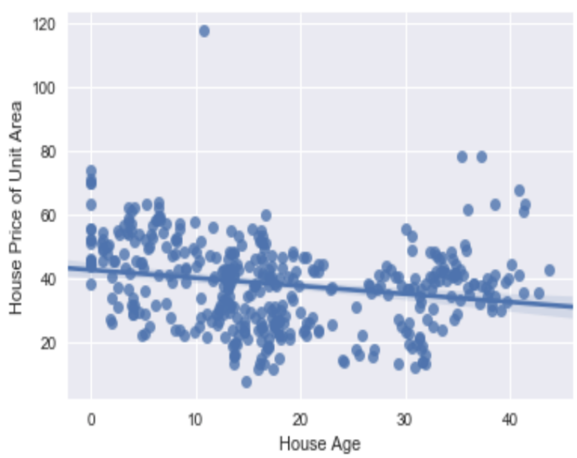

Figure 2: Correlation between the age of a house and the house price of unit area.

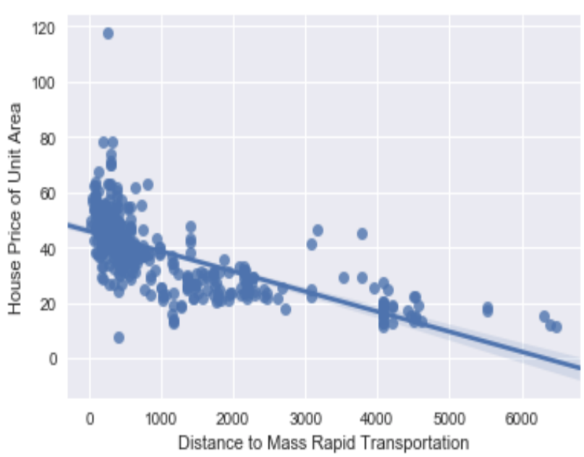

Figure 3: Correlation between the distance to mass rapid transportation and the house price of unit area.

I also found that the number of convenience stores in an area does not effect whether or not someone will buy a home. As we see below in Figure 4, there are quite a few people that bought houses with zero convenience stores in the area.

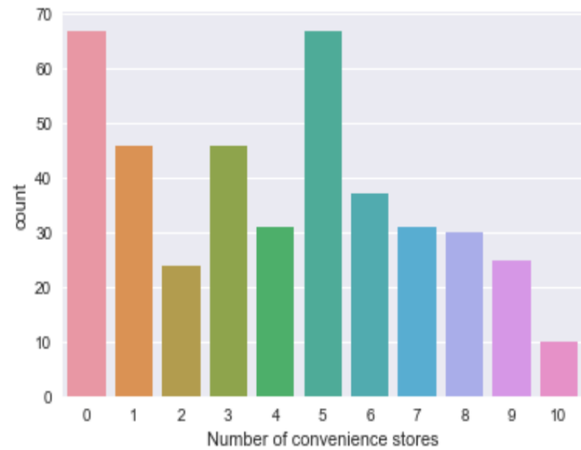

Figure 4: Number of houses bought at each amount of convenience stores.

5 DISCUSSION

The results from my research and visualizations allow me to understand the real estate industry a little better. The appraisal process of a property takes numerous factors into consideration before putting a price on the home. It takes into account everything from the average age of the neighborhood and family size to the proximity to recreational amenities and size of the house. These factors play a major role in not only the price and value of a home, but also who will likely buy a home. Most likely, if you have children and you want them to be able to play outdoors, you will move to an area where there might be a big backyard and several families in the neighborhood. It is the people’s desires that give value to real estate.

My data allowed me to see the numerous factors that contribute to a home's value and conclude that proximity to mass rapid transportation and the age of the house both negatively affect a home's price. It can be concluded from my data and visualizations, that generally speaking, you are likely to find cheaper real estate if further from mass rapid transportation or the house/apartment is older. Another conclusion that we can draw from the data is that the amount of convenience stores in an area is not a high priority on many people's criteria when buying a home. The visualization showed that the same number of people bought homes in the area with 4 convenience stores and 7 convenience stores. There was not a very strong trend to conclude that the number of convenience stores had an impact on whether or not someone bought a house. The visualizations and data like this help people understand real estate valuations and trends over the years.

6 CONCLUSION

My data set may not have been fully reflective of real estate valuation factors because it was data taken the early 2010s. However, we can gather rather general conclusions from the research and visualizations made. The real estate industry is a complex mix of multiple industries—a major intersection being the involvement of banks. Without these industries, there would be a major issue in the housing market. The mix of industries are crucial parts, but the most important factor in making the real estate industry hold any value are the people. It is human's needs and desires in real estate that makes the whole industry have value. Using a dataset that kept a record of real estate transactions and different factors about each, I was able to answer my research questions: Does the age of a home or the distance to mass rapid transportation affect the price of the home? And, does the number of convenience stores in the area affect the amount of people that will purchase a home? Looking at real estate transactions in this way, I'm more inclined to believe that the real estate industry is a more complex market than previously thought. It has a major influence on not only our nation's economic health, but the global economy as well and it continues to remain an important part of everyone's life.

REFERENCES

[1] Education Committee. 1951. "The Appraisal of Real Estate". *American Institute of Real Estate Appraisers*. Chicago, IL.
[2] Holt, C. "Banks and Real Estate Prices". *Swiss National Bank*.
[3] Smith, L. 2018. "What Is a Housing Bubble?"*SmartAsset*. https://smartasset.com/mortgage/what-is-a-housing-bubble
[4] Amadeo, K. 2016. "Stock Market Crash of 2008". https://02038.com/wp-content/uploads/2017/03/Stock-Market-Crash-of-2008.pdf
[5] Grether, David M. and Mieszkowski, Peter. 1974. "Determinants of real estate values." *Journal of Urban Economics*, 1 (2). pp. 127-145. https://authors.library.caltech.edu/83731/
[6] Hammer, T.R., Coughlin, R.E. and Horn, E.T. 1974. "The Effect of a Large Urban Park on Real Estate Value." *Journal of the American Institute of Planners*. Vol 40. p. 274-277
[7] Haurin, D. and Brasington, D. 1996. "School Quality and Real House Prices: Inter-and-Intrametropolitan Effects." *Journal of Housing Economics*. Vol. 5. pg. 351-368
[8] Carroll, T.M., Clauretie, T.M., and Jensen, J. "Living next to godliness: Residential property values and churches." *The Journal of Real Estate Finance and Economics*, Vol 12, 319-330
[9] Casselman, B. 2018. "Housing Market Slows, as Rising Prices Outpace Wages." *The New York Times*
[10] Haider, M and Miller, E. 2018. "Effects on Transportation Infrastructure and Location on Residential Real Estate Values" *The National Academies of Sciences Engineering Math*