

Xi Chen

Professor Barry Waite

PPD 631: Geographic Information Systems for Policy, Planning and Development Fall, 2019

Analysis of Airbnb Housing on Rental Pricing in Los Angeles County

1.Introduction:

The basic economic principles of supply and demand apply to all kinds of different markets, including both the hotel industry and the rental housing market. Supporters for the sharing economy insist that the sharing economy on the housing market, in other words, Airbnb, allows residents to earn extra income and decrease the unnecessary costs brought by the agents before (Horn & Merante, 2017). However, opponents of Airbnb claim that the existence of Airbnb is highly likely to decrease the supply in the long-term rental housing units and cause a remarkable increase in rental pricing as a result.

In short, Airbnb provides a platform for landlords to enter into the tourist market by converting traditional residential homes into hotel rooms for tourists. While Airbnb facilitates more tourism growth, it distorts the rental housing market by reducing the supply of affordable housing with the allure of substantial profits brought by such short-term rentals. New York has passed the legislation to compete against the potential impact of Airbnb on the rental pricing in 2018, which could be considered as the evidence for the rental pricing increase caused by Airbnb.

This project is focused on the relation between the existence of Airbnb and the pricing of rental housing in Los Angeles County. As is mentioned before, the crisis of housing affordability in Los Angeles have been intensified as Airbnb enters this city. From my perspective, there are chances that Airbnb distorts the rental housing market by reducing the supply of affordable housing remarkably. Thus, I would like to utilize GIS and other analytical tools to verify my hypothesis. First and foremost, I would examine the density of Airbnb, which is equal to the number of Airbnb listings in one neighborhood divided by the total housing units in that region. It would be helpful for me to see whether Airbnb-dense neighborhoods overlap the areas with high rental prices. Then, I would build a regression model between median rents and other

independent variables, including Airbnb listings, household income, crime incidents, population, and the number of restaurants and cafes in the accordingly neighborhoods. By including other relevant independent variables, the regression model could offer us insights into how the Airbnb listings exactly influence rents with other factors controlled.

2. Methodology:

2.1 Data Illustration:

All the data mentioned in this project is aggregated on the zip code level in Los Angeles County.

1) Median gross rent:

A. Illustration:

Median gross rent offers information on the average housing expenses per month for individuals. Specifically, it includes the contract rents as well as “the average monthly cost of utilities (electricity, gas, water, and sewer) and fuel (oil, coal, kerosene, wood, etc.), if these fees are paid by renters, according to U.S. Census Bureau, American Community Survey (ACS). The costs of water, sewer, and fuels are estimated yearly but transformed into monthly figures to combine the data with the monthly rents easily later. Gross rents can effectively normalize the variability while not all the renters have to pay for the utilities.

B. Reference:

United States Census Bureau. (n.d.). GROSS RENT Universe: Renter-occupied housing units more information 2013-2017 American Community Survey 5-Year Estimates. Retrieved from:

https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_B25063&prodType=table

2) The Number of Airbnb listings:

A. Illustration:

The number of Airbnb listings refers to the number of rental housing in the platform of Airbnb. In other words, it demonstrates the supply of housing in Airbnb quantitatively.

B. Reference:

Zillow. (n.d.). Housing Data. Retrieved from: <https://www.zillow.com/research/data/>

3) Median household income:

A. Illustration:

Median household income is often utilized to describe certain household's economic status or help grasp the overall trend of economic development in the whole country. It refers to the total amounts of income from all residents whose ages are above 15 in one house.

B. Reference:

United States Census Bureau. (n.d.). SELECTED ECONOMIC CHARACTERISTICS, 2013-2017 American Community Survey 5-Year Estimates.

Retrieved from:

https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_DP03&prodType=table

4) Crime incidents:

A. Illustration:

The crime incidents are from Los Angeles County Sheriff's Department and include 30 types of crimes occurring in the jurisdiction of the Los Angeles County Sheriff's Department.

B. Reference:

Los Angeles County Sheriff's Department. (n.d.). Part I and II Crimes. Retrieved from: <http://shq.lasdnews.net/CrimeStats/CAASS/desc.html>

5) Miscellaneous:

A. Illustration:

There are other independent variables such as populations, public transportation, the number of restaurants and cafes in one neighborhood. In the part of public transportation, there will be only two types of values: 1 and 0. Specifically, if there is any metro or bus station in the zip code area, the value of public transportation for this zip code area would be 1. Instead, the value would be 0.

B. Reference:

a) Population:

United States Census Bureau. (n.d.). American Community Survey 5-Year Estimates, Retrieved from:

https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_DP03&prodType=table

b) Public Transportation:

The Neighborhood Data for Social Change (n.d.). Public Transit Use (LA). Retrieved from: <https://usc.data.socrata.com/Los-Angeles/Public-Transit-Use-LA-/ff5n-m9wp>

c) Restaurants and cafes count:

Yelp Dataset. (n.d.). Yelp Dataset_Business. Retrieved from: <https://www.yelp.com/dataset>

2.2 Data Limitation:

For one thing, there exists much missing data which could damage the validity of this project to some extent. In order to attain as more data of all related variables in the zip code level of Los Angeles County in recent years as possible, I have to utilize the data from 2015-2016 to make the analysis. However, since the analysis is based on the data three years ago, it might have limited explanation power for the current situation.

For another, the independent variables are insufficient. It is reasonable to say that there are still other independent variables that might affect the prices in the long-term housing market, which suggests that the regression model might be inaccurate under certain circumstances.

2.3 Data Processing:

1) Missing data:

There are 509 observations in total from 2015 to 2016 on the zip code level in Los Angeles County. By excluding observations that contain null values via python, there are 407 valid observations left.

Table 1: Missing Data in Variables

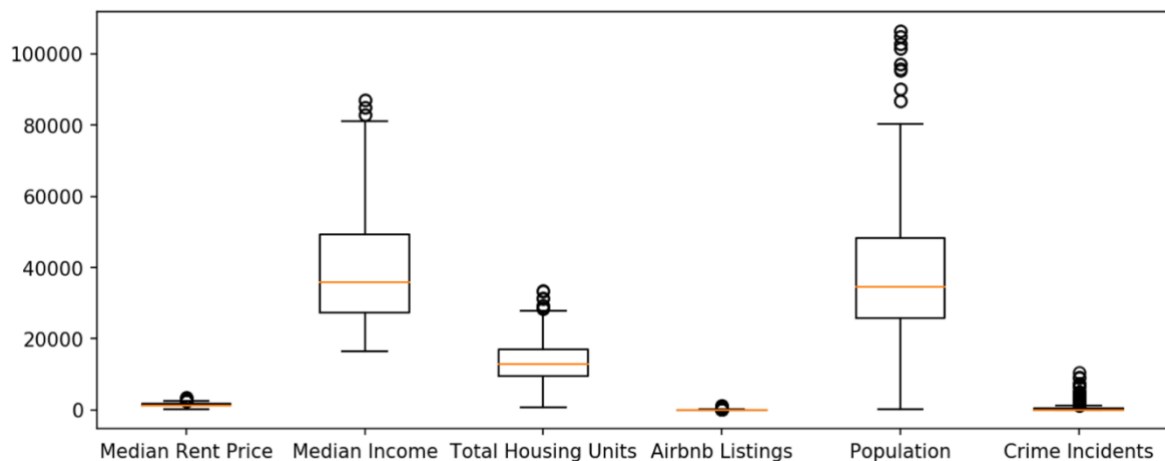
Total (Original Data)	509	
Missing Data	Median Gross Rent	11
	Median Household Income	13
	# of Airbnb Listings	89
	Population	4
	Crime Incidents	22
Useful Data	407	

2) Regression model:

To increase the goodness of fit for the regression model in this project, I use the formulation of ln for both the x and y variables (the number of Airbnb listings, median household income, crime incident, population, public transit, the number of restaurants/cafes and median gross rent).

3) Data outliers:

Figure 1: Distribution of Relevant Variables

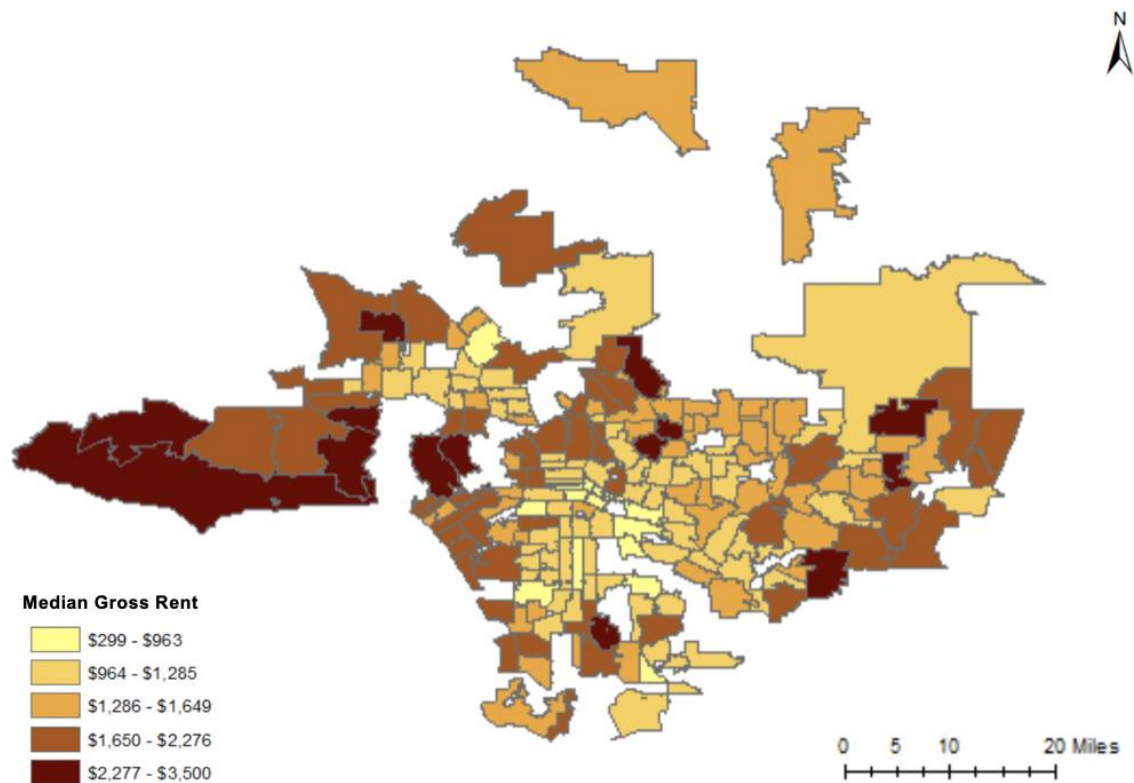


According to the boxplot above, there are several points outside whiskers, which are regarded as outliers. There are 42 observations considered as outliers among 407 observations in total. However, these outliers are kept, for I find that there is a slight difference between the models with outliers or not.

3. Analysis:

3.1 Median Gross Rent:

Figure 2: Median Gross Rent in Los Angeles County

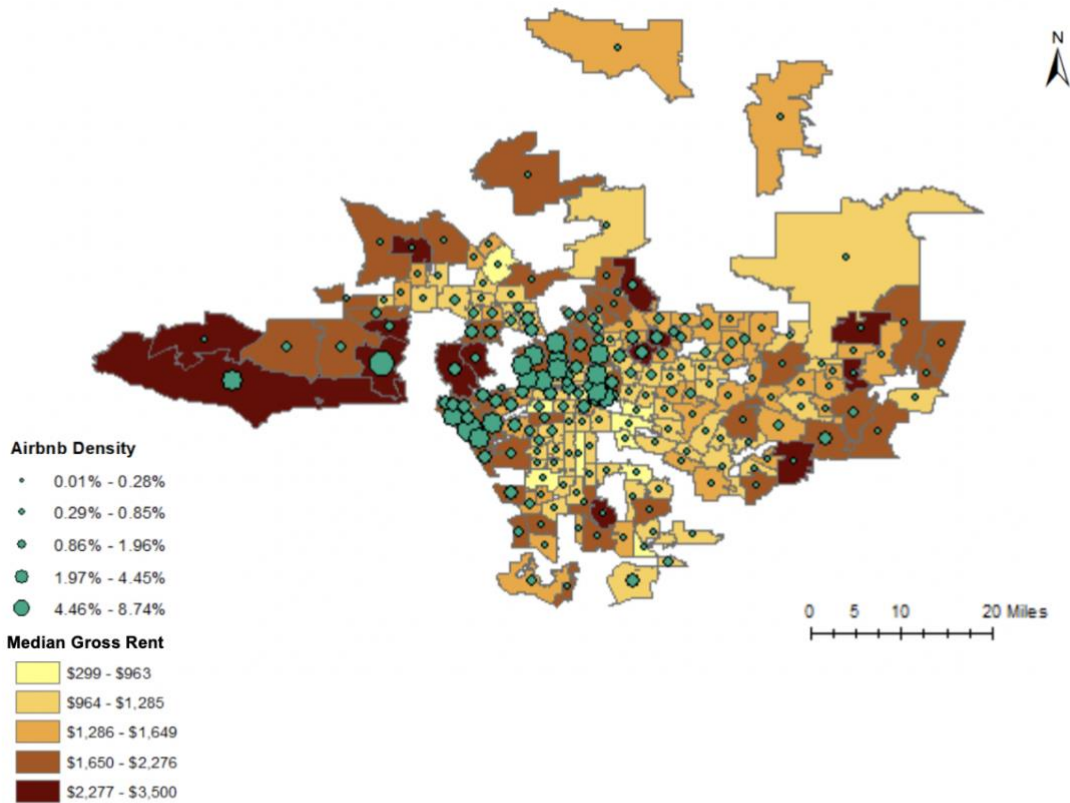


Source: GROSS RENT Universe: Renter-occupied housing units more information 2013-2017 American Community Survey 5-Year Estimates.
https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_B25063&prodType=table

As the map indicates, the edge areas of Los Angeles County generally have higher median gross rents than the center areas. Besides, the rents are remarkably high in the regions on the western side, such as Santa Monica, the Westside and so forth. The rents in the eastern regions have sharp variation and the rents in some areas in this region are terrifying as well. However, the rents in the center and northern areas are relatively low, where the median gross rents are possible to be below \$1000.

3.2 Median Gross Rent & Airbnb Density

Figure 3: Median Gross Rent VS Airbnb Density in Los Angeles County

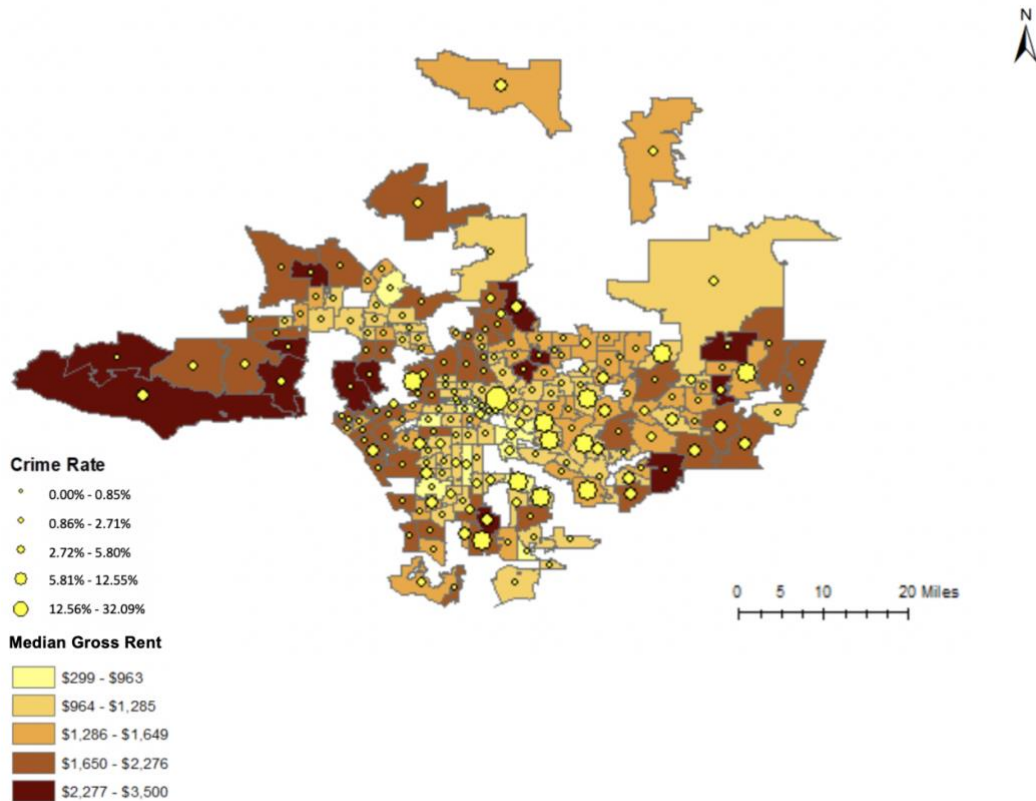


Source: CONTRACT RENT
Universe: Renter-occupied housing units
2013-2017 American Community Survey 5-Year Estimates
https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_B25056&prodType=table

The number of Airbnb listing refers to the number of housing in the Airbnb platform, and however, it could be influenced by the total number of housing units in the area. In order to normalize the variability, I divide the number of Airbnb listings by the total number of housing units to calculate the Airbnb density. From this map, there are some areas which have both high Airbnb density and high rental prices. Nevertheless, the center areas where the rental pricing is relatively low are Airbnb-dense. Thus, there seems some relation between Airbnb distribution and rental housing pricing based on the map but we cannot draw an accurate conclusion.

3.3 Median Gross Rent & Crime Rate

Figure 4: Median Gross Rent VS Crime Rate in Los Angeles County



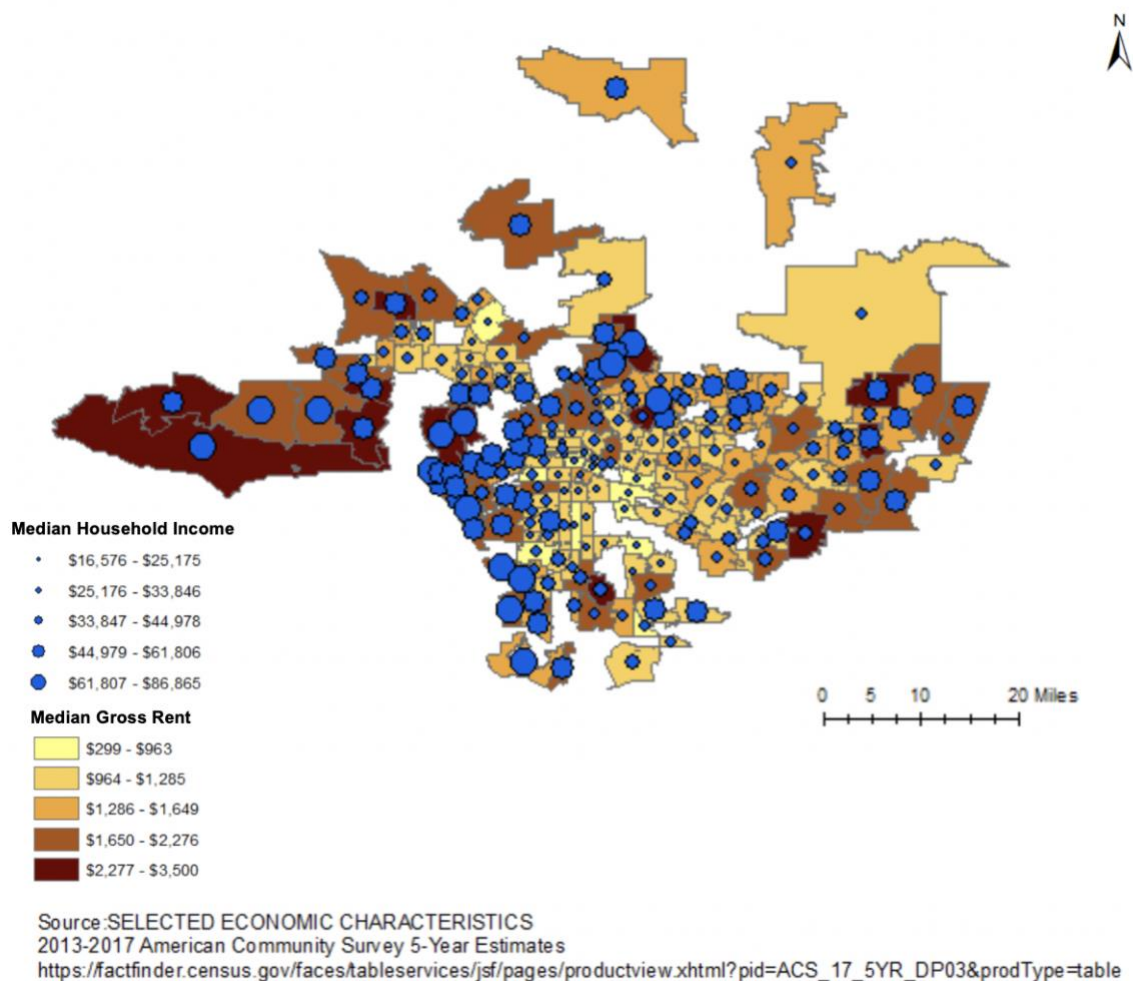
Source: Historical Crime data Part I & II in the Los Angeles County Sheriff's Department's jurisdiction
<http://shq.lasdnews.net/CrimeStats/CAASS/desc.html>

Security is a crucial factor which would affect residents' decisions when choosing where to live. Thus, the above map attempts to explore the relation between crime incidents and median gross rent. However, since there are some areas with more crime incidents than others due to their relatively larger population, crime rate is calculated to avoid such bias. The crime rate from the above map is equal to the number of crime incidents divided by the total number of populations in the accordingly area. As the map above suggests, the lower the crime rate, the higher the median gross rent. For instance, with relatively higher rents, the western areas enjoy pretty low crime rates as well, which are below 2.5% on average. Nevertheless, the center areas have low median gross rents and their high crime rates could be above 12% in the meanwhile. Due to the calculation of crime rate, there seems to exist a skewing of the data in areas with crowds in the daytime but few residents. However, even though the crimes in those areas might

be caused by people who don't live in the neighborhood, such aforementioned areas are insecure because of its too many crime incidents even with smaller population. The relatively high crime rate from calculation could be reasonable to some extent. Besides, when building the log-log regression, I replace the variable "crime rate" with two variables "population" and "crime". The potential skewing of data might be alleviated.

3.4 Median Gross Rent & Median Household Income

Figure 5: Median Gross Rent VS Median Household Income in Los Angeles County



From this map, it is clearly illustrated that the areas with high median household income are overlapped with those where the rental prices are high. For example, in the western area where the median gross rent could be higher than \$2277, the median household income is also far above average. The areas in the center of Los Angeles County have relatively lower median gross rent and they have median household income far below average as well.

3.5 Log-log Regression Model

Based on the maps above, it is reasonable to conclude that the variables including Airbnb listing, crime incidents, and median household income have relations with the turbulence of housing rental prices. However, it is still challenging for us to tell to what extent the existence of Airbnb influence on the rental housing prices, especially with other variables uncontrolled.

Therefore, I build a regression model by setting median gross rents as the outcome variables and others as the decision variables. Comparing different regression models, such as linear regression, regression with semi-log and regression with both X and Y log, I choose the log-log regression model considering its relatively higher coefficient of determination.

The formula is as follows:

$$\text{Ln (Rent)} = \beta_1 * \text{Ln (Median_Income)} + \beta_2 * \text{Ln (Num_Airbnb_Listings)} + \beta_3 * \text{Ln (Num_Crimes)} + \beta_4 * \text{Ln (Population)} + \beta_5 * \text{Public_Transit} + \beta_6 * \text{Ln (Num_Restaurants)} + \epsilon$$

Table 2: Regression Statistics

<i>REGRESSION STATISTICS</i>	
MULTIPLE R	0.915838202
R SQUARE	0.8387596122
ADJUSTED R SQUARE	0.8363410064
STANDARD ERROR	0.09919699336
OBSERVATIONS	407

Table 3: Coefficients of Log-log Regression Model

	Coefficients	Standard Error	T Stat	P-Value
Intercept	1.183670368	0.1634971959	7.239698281	0
Log Number of Airbnb Listings	0.01551670379	0.003371970065	4.601673056	0.000005635243925
Log Median Household Income	0.5454783888	0.01285567237	42.43094979	0
Log Crime Incidents	0.002927384575	0.002313507887	1.265344541	0.2064844302
Log Population	-0.005777963512	0.006975170293	-0.8283616412	0.4079597004
Public Transit	-0.0008005469545	0.01194812135	-0.06700191029	0.946613659
Log Restaurant/Café Count	0.005661261498	0.004962248469	1.140866189	0.254608196

According to the tables above, a log-log regression model is established based on 407 observations. The R square, or the efficient of determination, is around 0.84, which suggests that this model enjoys relatively excellent goodness of fit. Thus, it is rational to conclude that the model to illustrate the correlation between the number of Airbnb listings and the median gross rent in certain area on the zip code level in Los Angeles County. Besides, comparing the coefficients of all the independent variables, the median household income has the most significate influence on median gross rents than other variables while the number of Airbnb listings enjoys the second most obvious correlation with median gross rents. Their extremely small P-values suggests the validity of the aforementioned assumptions.

Thus, according to the regression model, although we cannot prove there exists a causal relation between the number of Airbnb listings and median household income, it is reasonable to conclude that the number of Airbnb listings has positive correlation with the median gross rent.

4. Conclusion:

According to the aforementioned analysis, first and foremost, I explore the relations between Airbnb density, crime rate, median household income, and median gross rents. Then, to figure out the correlation between the number of Airbnb listings and median gross rents, I build a log-log regression model to prove that the positive correlation between them on the zip code level in Los Angeles County exists with other variables controlled, however, which cannot imply causation necessarily.

Above all, the existence of Airbnb makes contribute to increase in the rental housing prices to some extent and is likely to deteriorate the housing unaffordability in Los Angeles County. Compared with the entrance of Airbnb, median household income is obviously positively correlated with the growth of rental prices in Los Angeles County in recent years.

Therefore, when making public policies to deal with issues of rapid growth in rental housing prices, the policymakers should take into serious consideration the potential impact exerted by the entrance of Airbnb on rents, considering the potential positive correlation between rents and Airbnb listings. Considering the rigorous tourism in Los Angeles County, the effects from Airbnb existence might be even more remarkable and the policymakers should implement some measures, such as legislation, regulation, taxation and other alternatives, to limit the possible negative effects of Airbnb on the rental housing market in the future.

5. Reference:

Horn, K. & Merante, M. (2017). Is home sharing driving up rents? Evidence from Airbnb in Boston [Electronic version]. *Journal of Housing Economics*, 38, 14-24

Zillow. (n.d.). Housing Data. Retrieved from: <https://www.zillow.com/research/data/>

United States Census Bureau. (n.d.). GROSS RENT Universe: Renter-occupied housing units more information 2013-2017 American Community Survey 5-Year Estimates. Retrieved from:

https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5_YR_B25063&prodType=table

United States Census Bureau. (n.d.). SELECTED ECONOMIC CHARACTERISTICS, 2013-2017. Retrieved from:

https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5_YR_DP03&prodType=table

United States Census Bureau. (n.d.). American Community Survey 5-Year Estimates, Retrieved from:

https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5_YR_DP03&prodType=table

Los Angeles County Sheriff's Department. (n.d.). *Part I and II Crimes*. Retrieved from:

<http://shq.lasdnews.net/CrimeStats/CAASS/desc.html>

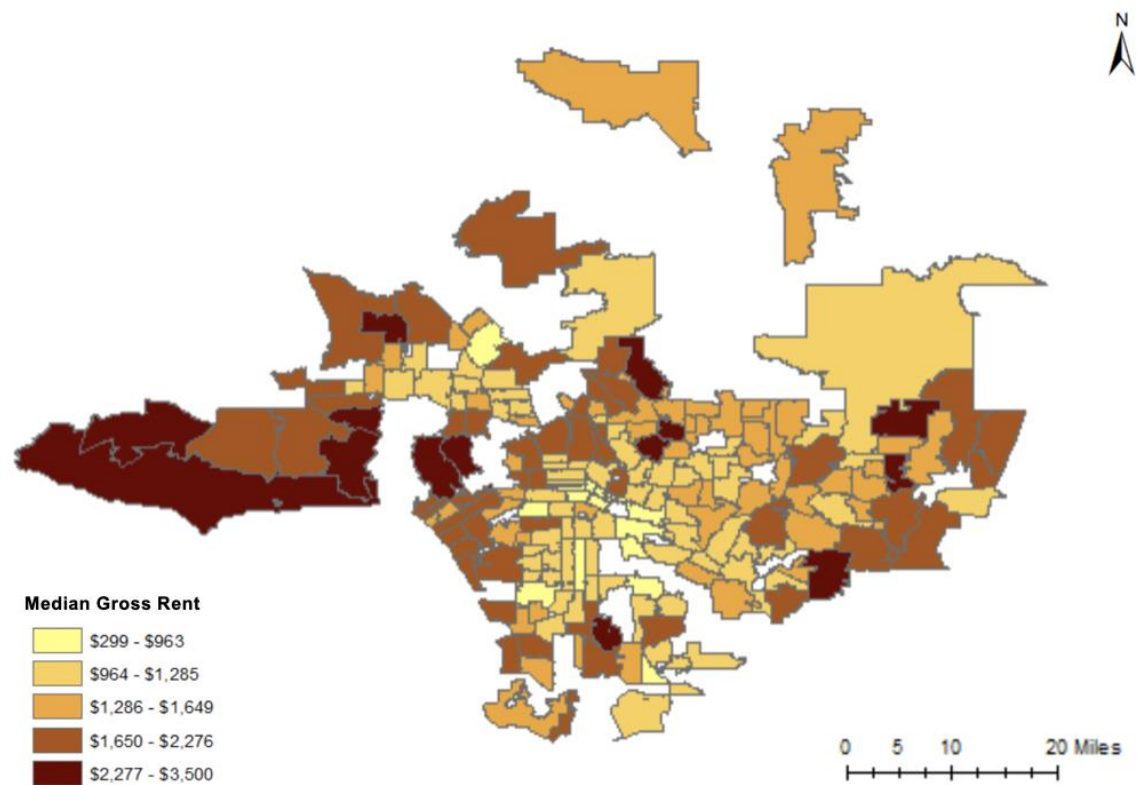
The Neighborhood Data for Social Change (n.d.). Public Transit Use (LA). Retrieved from:

<https://usc.data.socrata.com/Los-Angeles/Public-Transit-Use-LA-/ff5n-m9wp>

Yelp Dataset. (n.d.). Yelp Dataset_Business. Retrieved from: <https://www.yelp.com/dataset>

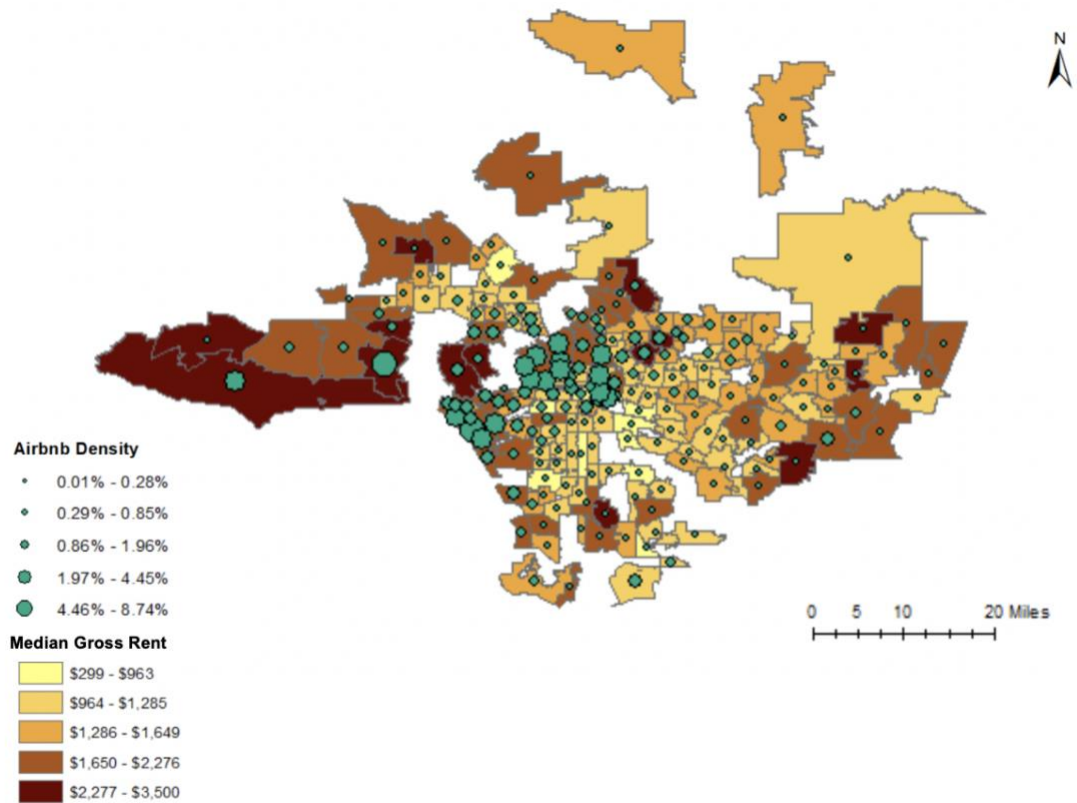
6. Appendix:

Figure 1: Median Gross Rent in Los Angeles Count



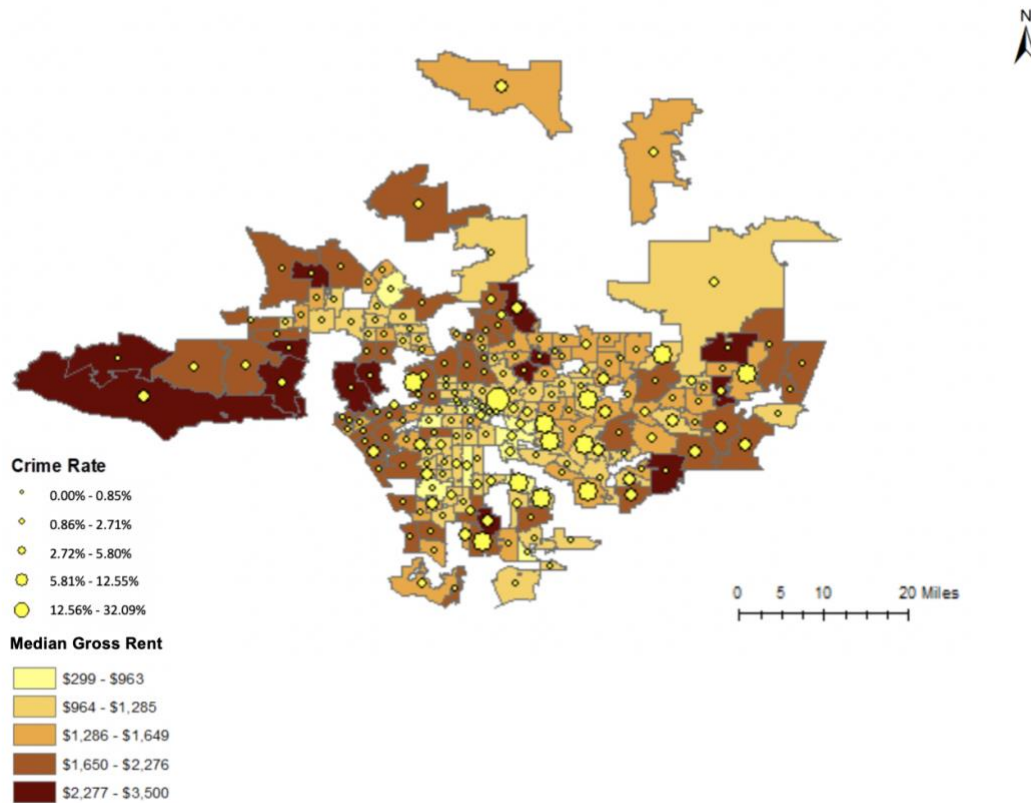
Source: GROSS RENT Universe: Renter-occupied housing units more information 2013-2017 American Community Survey 5-Year Estimates.
https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_B25063&prodType=table

Figure 2: Median Gross Rent VS Airbnb Density in Los Angeles Count



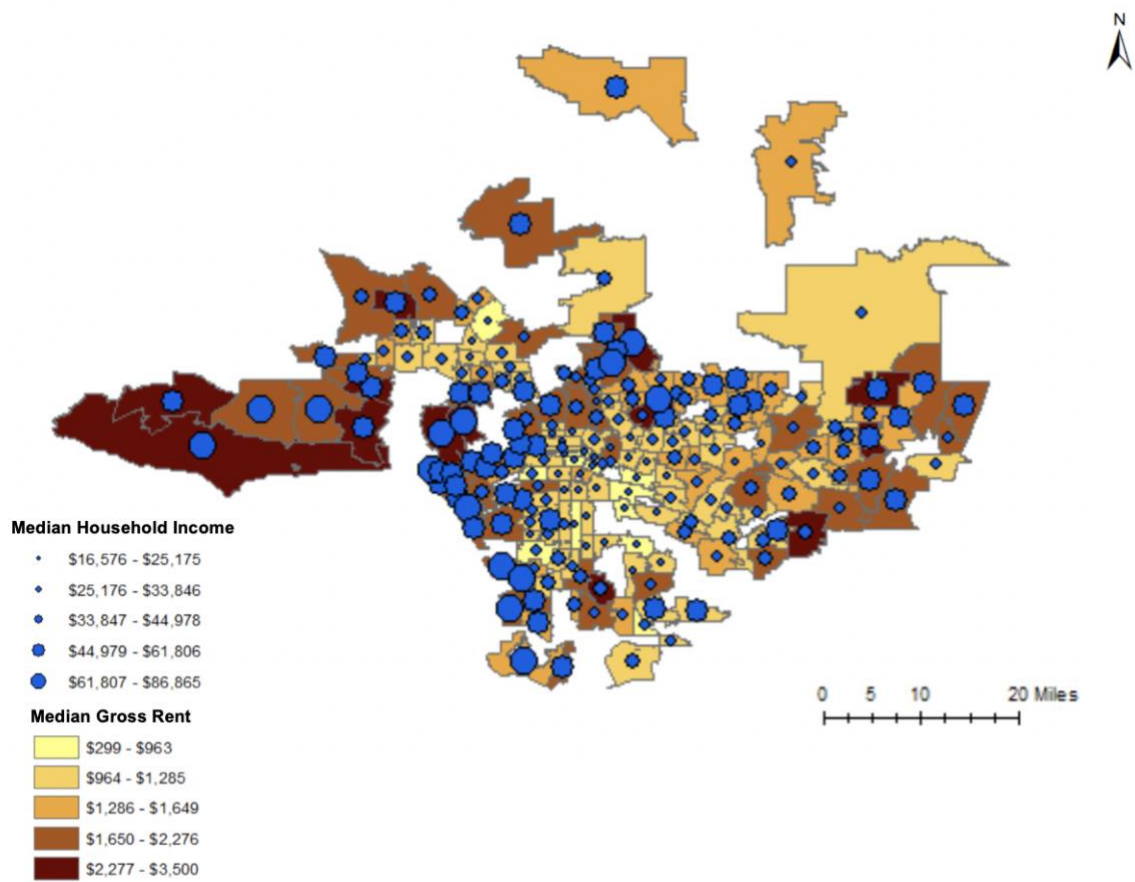
Source: CONTRACT RENT
Universe: Renter-occupied housing units
2013-2017 American Community Survey 5-Year Estimates
https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_B25056&prodType=table

Figure 3: Median Gross Rent VS Crime Rate in Los Angeles Count



Source: Historical Crime data Part I & II in in the Los Angeles County Sheriff's Department's ; jurisdiction
<http://shq.lasdnews.net/CrimeStats/CAASS/desc.html>

Figure 4: Median Gross Rent VS Median Household Income in Los Angeles Count



Source: SELECTED ECONOMIC CHARACTERISTICS
2013-2017 American Community Survey 5-Year Estimates
https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_DP03&prodType=table