Relationship between crime and sharing economy: Impact of Airbnb listings on the crime in Philadelphia

Abstract

The recent fast growth of sharing economy has attracted attentions of many studies. As the leading platform in accommodation and lodging services, Airbnb and its influence in neighborhood is important for urban policy making. In this paper, we study whether the introducing of Airbnb to neighborhoods would affect local safety, which is an important indicator of the life quality. We focus on Philadelphia as an example, with data of more than 6 thousand Airbnb listings and 10 year's crime data. Statistical methods, such as regression discontinuity analysis, are applied on the block group level. Our analyses show that Airbnb does not affect crime in a pre-specified neighborhood of the unit.

1 Introduction

Sharing economy, also known as peer-to-peer economy, has been exponentially gaining importance during the past decade. The fundamental concept of this economy is to make use of the underlying services and commodities via fee-based transactions. The sharing economy has seen an enormous growth during the past few years owing to the development of information technology and specifically the growth of online platforms which play as a new form of mediator between the provider and the consumer Kaplan and Haenlein [2010]. Airbnb, is one of the well-known platforms of this kind offering accommodation and lodging services to the consumers by connecting the home owners to the potential consumers via an online platform.

Since its emergence in 2008, Airbnb has marked a phenomenal growth and an increasing popularity. The Economist named Airbnb as 'the most prominent example" of sharing economy in 2013 Economist [2013]. In a report published on 2015, Airbnb claimed that it has served over 50 million guests with a market value of $24 billion and expected revenue of $10 billion by 2015 Winkler and Macmillan [2015]. by 2016, Airbnb claims that it has reached out to over 191 countries, 34,000 cities and more than 2 million listings worldwide Airbnb [2016a]. To put that number into perspective, the largest chain hotel, the InterContinental Hotel groups PlanetDen [2015], have branches in only 100 countries and 5070 locations worldwide Group [2015]. Today, Airbnb has considerably affected the urban landscapes. In New York for example, roughly 5% of the vacant apartments on average have turned to Airbnb units, leaving alone those units where the hosts rent only a section of their residents (only 10% of the units in New York are whole apartments). This number increases to over 23% of the vacant apartments in some neighborhoods such as West Village and Williamsburg Airbnb [2015] while decreasing the vacancy rates and dramatically increasing the tourist agglomeration in some neighborhoods.

Same as every other big economy, it's expected that this new phenomenon come with certain social consequences. Although the literature around this issue is scarce in part due to being a nascent phenomenon, some studies have reports have highlighted some of these consequences. Among the most important ones are influencing the housing market, providing additional incomes for urban communities, expanding the tourism experience to a greater number of neighborhoods in cities and revitalizing the otherwise isolated neighborhoods. While most of the positive impacts are claimed by Airbnb, no data-driven research has been presented so far to investigate the social consequences of this new urban landscape (i.e. Airbnb listings) in the neighborhoods.

In this study we will attempt to discover the social impacts of Airbnb on the neighborhoods by using crime as an indicator of the neighborhood quality. This research intends to investigate the extent to which the emergence of the Airbnb units in neighborhoods affect the occurrence of crime. Two datasets were used to answer this question: first, the Airbnb data has been scraped from the Airbnb website for the city of Philadelphia. The data includes up to 130 variables for each Airbnb unit including information on the nightly price, additional guest costs, deposit, neighborhood, amenities, room type etc. Additionally, this dataset includes the reviews of the clienteles for each Airbnb unit as well as the date of the reviews and the user ID. Second, the crime data for the city of Philadelphia covering 32 types of crime occurred in this city from 2006 up to 2016. This data has been provided by OpenDataPhily (see `https://www.opendataphilly.org/`) and includes data on the location, crime type and the hour and the date of the crime. This research employs spatiotemporal analysis methods to study the possible changes in the crime rate in the proximity of the Airbnb units by investigating the occurrences of different crimes in each proximity within the past ten years. It is expected that our research informs the policy makers, urban planners and the communities of both positive and negative impacts of Airbnb on safety, as an important indicator of the life quality.

2 Background

2.1 The impacts of Airbnb on urban communities

The influences of Airbnb on the qualities of neighborhoods have not yet been studied deeply in part due to the fact that this phenomenon has not been introduced to many cities long enough. Not surprisingly, most of the studies conducted by Airbnb claim that Airbnb has enhanced the life quality and the tourism experience. Airbnb claims that it plays a key role in revitalizing the neighborhoods by expanding the spatial domain of tourism. Roughly, 70 percent of the Airbnb units are outside of the traditional touristy areas Airbnb [2016b]. In New York, for example, 82% of the Airbnb listing are located out of midtown Manhattan generating a total of $104 million in one year outside of Manhattan Airbnb [2016b]. In better words, the income from the tourism industry is now shared with peripheral areas and the local population that were deprived from this industry before Airbnb.

Another claim by Airbnb is that it has ameliorated the wealth distribution and empowered the weaker communities. According to the Airbnb polls more than 50% of Airbnb owners are medium to low income households and 53% of the owners claimed that this business helped them to stay in their homes while 48% claimed that the income from renting their living space helped them to pay regular household expenses such as groceries and rent. It is important to note that the entire neighborhood is likely to benefit from the influx of tourists since 42% of the spending is in the neighborhoods where the guests stayed at Airbnb [2016b]. Other than the direct economic effects, it is argued by some commentators that this type of business is likely to improve the social capital in such communities. Since Airbnb is essentially based on peer-to-peer communication, it is a first step towards generating new ties to the outside community and hence improving the social capital of the host and the guest for that matter Bialski [2012].

Notwithstanding the growing popularity, Airbnb has received considerable critiques from commentators in different fields. Some scholars believe that Airbnb units imposes negative impacts on the regular rental markets since Airbnb rentals are more profitable and this issue minimizes the chance of finding affordable housing for locals Southern California Public Radio [2015], James Ball and Franklin [2014], Lee [2016] this issue is more tangible for cities where lack of affordable housing is already troublesome Lee [2016] and those neighborhoods where the density of Airbnb listings is higher Kudler [2015]. In Los Angeles alone, the average monthly rents have increased by $6 over the past 5 years due to the increasing impact of Airbnb Southern California Public Radio [2015].

The rising rents for neighborhoods, as a result of the increasing number of Airbnb listings, comes with certain socioeconomic consequences. First, the lower-income will be replaced by the medium and higher-income which in part leads to fewer access to institutions and public services for the lower-income Lee [2016], Kennedy and Leonard [2001]. Also, given that the shared economy is based on mutual trust, it is also highly affected by racial discrimination. According to a new study African American home owners earn 12% less Edelman and Luca [2014], Todisco [2015]. This issue becomes more important in disadvantaged neighborhoods that are mostly occupied by the minorities.

Among more direct impacts of the Airbnb listings are the presence of tourists in neighborhoods that are not normally noticed by tourists. While this issue is championed as an improvement in Airbnb advertisements, some commentators believe that the excessive presence of tourists in all neighborhoods kills the local spirit of these areas eviction Mapping Project [2014]. Some scholars argue that the presence of the tourists directly disturb the hosts' neighbors and negatively affect the public safety Morris [2015]. While the negative effects of Airbnb in neighborhoods have been discussed by a number of studies, very few of them have supported their arguments through comprehensive data analysis processes.

2.2 The spatial dimension of crime

The past few decades has seen a considerable number of theories highlighting the spatial patterns of crime in part due to the fact that crime data has been long available to scholars through the governmental data sources. Reviewing these theories is an important step to identify the possible crime-related impacts of the Airbnb listings as a new urban landscape with both commercial and residential functions.

A large body of literature on crime focus on the association of different crime types with commonly known demographic factors such as income, age, race, density, youth concentration, poverty, family conditions and economic inequality Roncek [1981], Greenberg and Rohe [1984], Byrne [1986], Sampson [1986], Roncek and Pravatiner [1989], Sampson and Groves [1989], Swartz [2000]. A number of theories go beyond the demographic factors and argue that the characteristics of the built environment are influential factors. Among the earliest theories of this line is the "defensible space" Platt and Newman [1975] which asserts that the aptly designed spaces increase social interaction among the inhabitants and improve the informal guardianship in the neighborhoods and this, in part, reduces the crime rate. In this theory the built environment takes on an active role in determining the crime rate Platt and Newman [1975], Jacobs [1961], Rosenbaum [1988], Perkins et al. [1993]. Based on this theory, a number of studies in the past decade use the space syntax techniques to analyze the impact of configuration of spaces on defensibility Chih-Feng Shu et al. [2000], Shu and Huang [2003], Hillier and Shu [2000], Jones and Fanek [1997].. Space syntax, primarily developed by Hillier and Hanson in late 60s offers a number of theories and techniques used to assess the space use Hillier and Hanson [1989]. For example, a study on the residential neighborhoods in Taiwan found that there is a strong correlation between the global integration of streets (i.e. a space syntax measure for accessibility) and low crime rates in low-income neighborhoods Shu and Huang [2003].

While the defensible space theory mostly focuses on the configuration of urban spaces, the "broken windows" theory argues that the appearance of the neighborhood is also an important factor Skogan et al. [1999]. According to this theory, criminals are more likely to commit crimes where they find the residents to be indifferent from or detached from their neighborhoods. Buildings with broken windows, trash-filled lots and vacant buildings are signatures of neighborhoods where social control is lacking. Scholars believe that increasing visibility in these neighborhoods can improve the levels of social control Trojanowicz and Baldwin [1982], Wilson and Kelling [1982], Kelling [1985], Greene and Taylor [1988], Greene and Mastrofski [1988]. The "routine activities" theory, on the other hand, argues that the demographic characteristic of the neighborhood, while being influential is not the most important, but rather primary hindrance against crime is direct and constant surveillance. All the three major theories discussed here consider the presence of people as an important factor for reducing crime in the neighborhoods.

Most of the studies discussed above aggregate the crime occurrences on the common census units such as blocks, block groups, tracts and zip codes Roncek [1981], Roncek and Pravatiner [1989], Gottfredson et al. [1991], Perkins et al. [1993], Loukaitou-Sideris et al. [2001]. Some studies argue that crimes are likely to cluster in crime prone neighborhoods. These studies use the "distance decay model" in their analysis which assumes that the crimes are most likely in certain locale and it diminishes as the Euclidian distance from these centers increase Cohen and Felson [1979], Bottoms and Wiles [1997]. Later, a considerable number of spatial studies took this model one step further and included the component of time in their analysis by differentiating between permanent hotspots and temporary ones known as ephemeral hotspots Sherman [1995], Block and Block [1995], Buerger et al. [1995], Gorr and Olligschlaeger [2002].

3 Data description

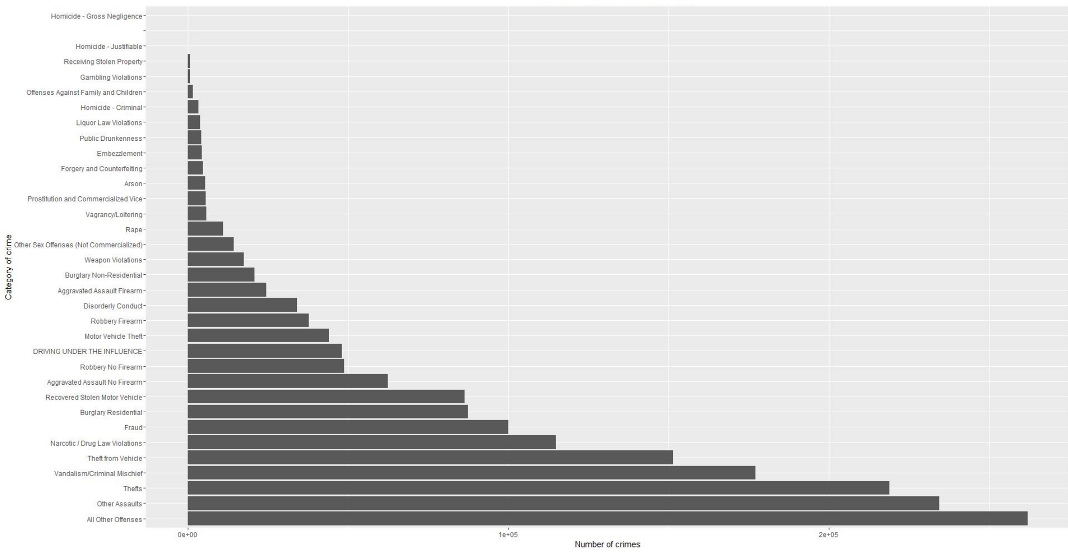
3.1 The crime data

The crime data includes 2140485 reported crimes in Philadelphia over a 10 years long period (i.e. from Jan 1, 2006 to August 18, 2016). This data entails the crime date and time, the latitude and longitude of the crime, the district where the crime has occurred in as well as 32 crime types. Not surprisingly the frequency of different crime types shows a high standard deviation (i.e. 94966.76) with "All other offenses" being the most frequent crime with 420920 occurrences and "Homicide - Gross Negligence" being the least one with only 12 occurrences.Among the more specific crime types, "Theft" is reported as one of the most frequent with 244544 occurrences. Figure 1 shows the frequency of all 32 crime types during the past 10 years in Philadelphia.

A quick investigation of the data reveals that the aggregated crime rate in different neighborhoods defer considerably with district 11 having the maximum number of crimes reported during the last 10 years and district 22 the minimum. A Kernel density map of the data helps to better understand the distribution of different types of crime. For example, Narcotic related crimes seem to be mostly concentrated in certain areas of the city while theft from vehicles are more homogeneously distributed. As we can see in the crime contour maps (Figure 2) narcotic related crimes are mostly focused in the northern side of the city whereas theft is mostly reported in the downtown and other crowded locations of the city and vehicle thefts are common place in most areas with a higher concentration in the center.

Since our research will include the analysis of Airbnb opening dates and the change of crime rate around the opening time, it is also important to investigate the temporal dynamics of crime in this city. Overall the crime rate in the majority of districts in Philadelphia shows a constant decrease since 2006. We can see in Figure 4 that

Figure 1: The frequency of different crime types in Philadelphia



generally the crime rate increases in summers and late spring. One reason for this increase could be the increased presence of people in the public spaces or increased rate of tourists in the city.

3.2 The Airbnb data

The Airbnb data are scraped from both the Airbnb website and their APIs. There are 3 major steps in scraping Airbnb data. The first step would be to scrape listing IDs from Airbnb website at the zip code level. The number of listings we can get from Airbnb API is limited to 20, while from the website is at most about 300, regardless of the geographical scale of the search. We thus scrape listing IDs from the website based on zip codes, in order to retrieve most of the listing IDs in Philadelphia. The second step is to scrape listing information through Airbnb API, which include almost all the information about the listing we can see on the website, such as listing type, number of beds, price, ratings, location, etc. The third step is to scrape all the reviews of the listing through Airbnb API. The reviews contain information such as reviewers' user

Figure 2: [left]Theft, [middle]Narcotics, [right]Vehicle theft

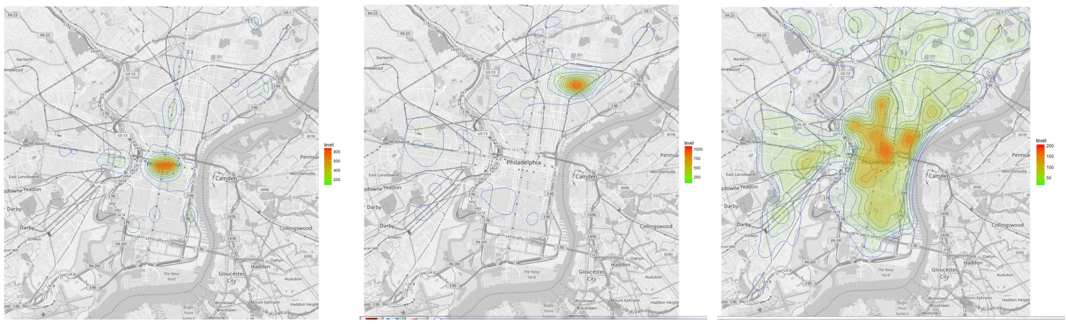


Figure 3: [left]Fraud, [middle]Assault, [right]Residential Burglary

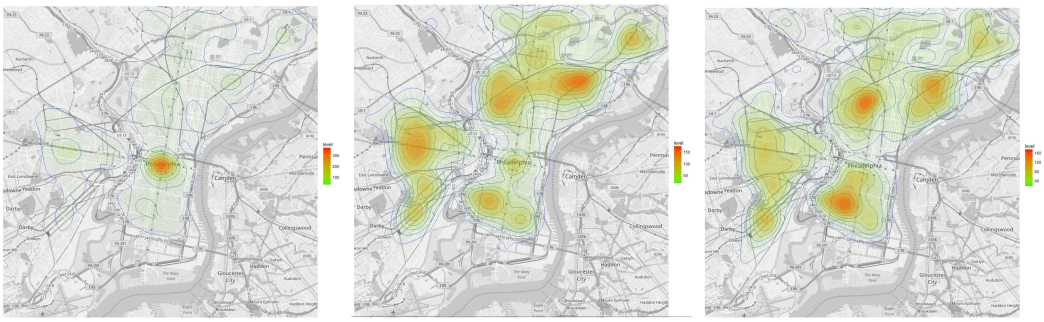
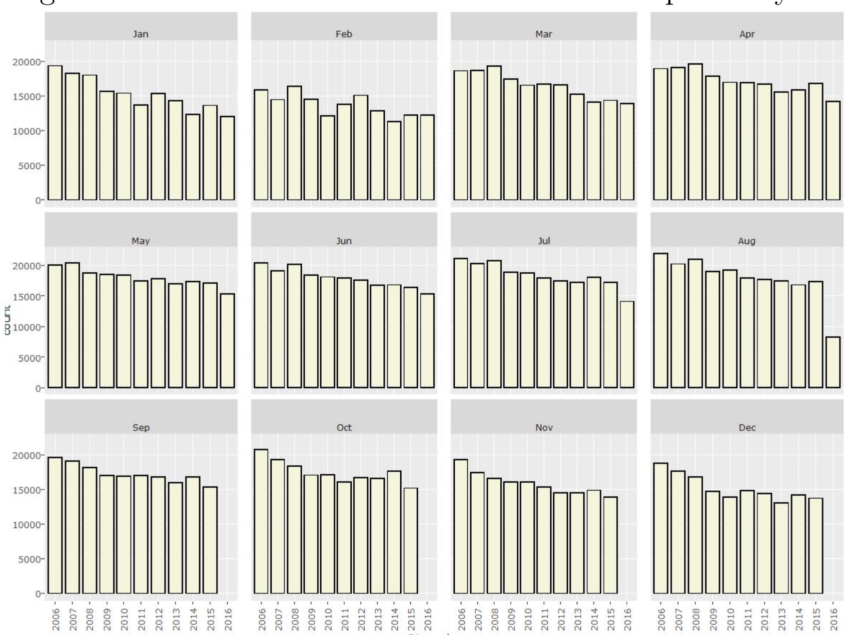


Figure 4: Crime rate in different months in the past 10 years



ID, review content, review date, etc.

We finally scraped 6,199 unique Airbnb listings in Philadelphia operated by 5,270 producers. Among the listings, 2,816 of them have reviews from 26,874 consumers, and the total number of reviews is 54,446. The consumers and reviews of each listing follows a power law distribution: most of the listings have few reviews, while some of them have many reviews. To note that although we scraped the Airbnb data based on zip codes, the data are still a reasonable representation of Airbnb: we got a proportion of duplicated listing IDs when scraping from different zip codes, which demonstrated that we have collected most of the available Airbnb listings; moreover, we have more listings compared to some other related data sets. The plots below summarize demographics of the city, as well as the Airbnb data and its spatial features.

Figure 5: Airbnbs within Philadelphia metropolitan area in red

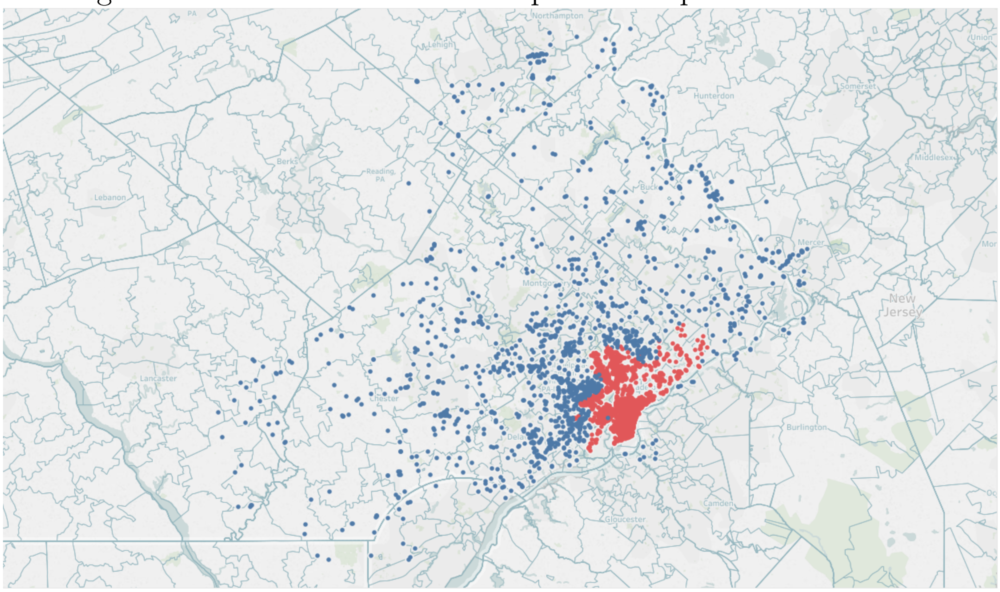
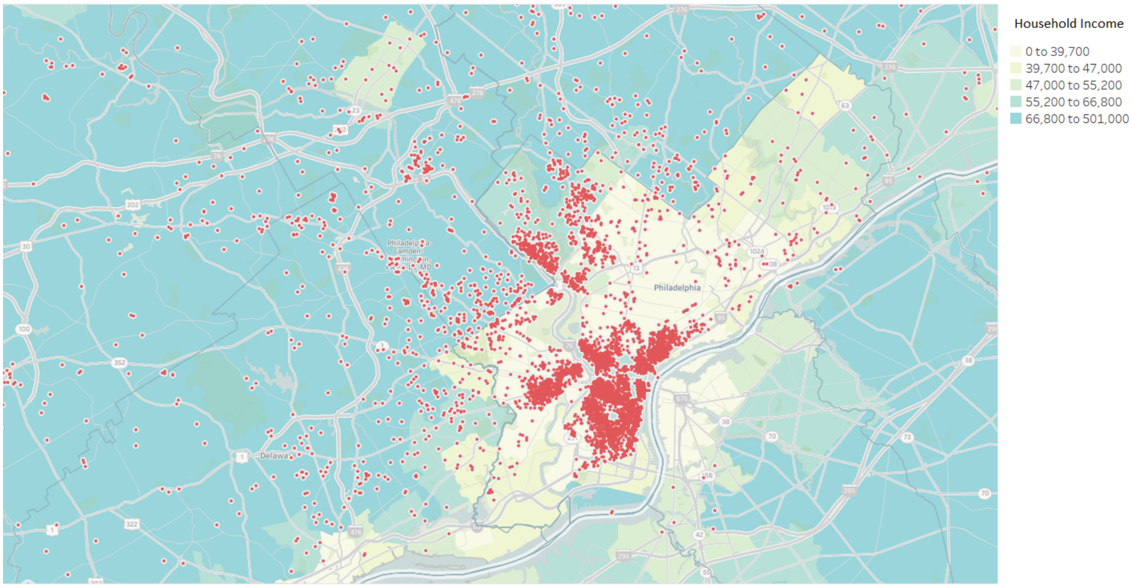


Figure 6: Median household income by zipcode



4 Methodology

4a Spatial Processing

Geocoding crime data. Although the major part of crime records have longitude and latitude information, about 16.5 thousand records are still missing coordinates. In order to spatially matching Airbnb listings and crimes nearby, we used Google Maps Geocoding API to geocode these recodes based on the addresses of where they happened.

Joining Airbnb listings and crime records spatially. In order to investigate the influence of Airbnb listings on its adjacency space, we calculated the number of different crime types in each year near each Airbnb listings. The adjacency are defined as a series of distances, including 50 m, 100 m, 250 m, 500 m, 1000m. We have such large number of listing and crime points that direct computing has high time complexity. In order to use spatial index to speed up the querying process, we used PostGIS, a spatial database plugin for PostgreSQL, to implement the spatial join process. The numbers of nearby crimes for each listing thus provide a series of variables for further analysis.

Aggregating Airbnb listings and crime records to block groups. Besides exploring the effects of individual Airbnb listings, how their existence effect neighborhoods can demonstrate our research topic from a different geographic scale. We choose block groups of Philadelphia as the spatial units. Airbnb listings are spatial joined to its corresponding block group. Crime records are aggregated into each block group for each month-year combination, and each crime type in order to have fine temporal granularity to see the effects of Airbnb listings. Such aggregation allows us to observe crime trends in a sufficiently aggregated time unit.

4b Statistical Analysis

We conduct regression discontinuity (RD) analysis to study if the introduction of a new Airbnb to a neighborhood (block group here) acts as a discontinuity, in turn affecting crime in the later stage. Addition of each new Airbnb to a neighborhood can incrementally affect crime. Such an effect is not temporally separable. Its spatial spreading is also difficult to account for, depending upon the definition of neighborhood. Therefore, we select the month-year of introduction of the very first unit in each neighborhood as time of discontinuity. We have data on reviews written for each Airbnb unit. The date of the first review is considered as a proxy for the date of introduction for that unit. Therefore month of introduction is the treatment. We aggregate monthly crime, and analyze values for three months before and after treatment. The date of treatment in overall sample of 191 block groups ranges from mid-2011 to mid-2016. We lose the maximum number of observations due to lack of sufficient post treatment data. We also lose observations since they cannot be matched to block groups in the city of Philadelphia. Post-subsetting and cleaning, the crime data does not have observable trend and the month-to-month variation reduces to negligible quantity. Therefore we do not de-trend or bin the crime data.

We also modify the categorization of crime. The initial 32 categories are collapsed into 14 categories, and some of them have been excluded for the lack of relevance to research question. For all analyses, consider the following types of crime; Assault, Burglary, Drunkenness, Fraud, Homicide, Narcotic, Offense, Other, Other Sex Offense, Prostitution, Rape, Robbery, Theft, and Vandalism. We also create a 'Total' crime column and conduct the same analyses.

Within the RD design, we fit several linear models and consider their performance. The first set of models assumes the quantity of crime to be a gaussian variable. Since it is a count variable in reality, we fit an additional set of models where the outcome takes Poisson distribution. In each case, the number of crime (of a given type or the total) is considered as the outcome variable. The standard explanatory variables included here here are time (-3, -2, -1, 0, 1, 2, 3) and a treatment indicator. Treatment indicator takes value 1 for time periods greater than or equal to 0. Of key interest is the main treatment effect. This effect is the quantity by which the outcome (crime) changes in shifting from control to treated status. We initially estimated an interaction effect between treatment and time. In theory, this would have allowed us to study the change in crime trend after treatment. However, due to the way these variables are constructed, estimating interaction would render the treatment main effect irrelevant. The main effect would then explain effect of treatment exactly when time = 0, which is the treatment period itself. Therefore we remove it from the specification. Finally, the significance of intercept indicates whether the average crime when both treatment and time is set to zero (exactly treatment month), is significantly different from zero.

It is important to note that this does not carry relevance in this analysis.

In the next set, we introduce block group fixed effects to the model. This specification is intended to control for effects specific to each neighborhood. It is possible that factors such as socio-economic makeup, public transport, access to tourist spots, and other amenities vary systematically for each of them. They can affect both, the introduction and usage of Airbnb, and crime. One coefficient is added for 190 out of 191 total census block groups considered in the analysis. A final modification to this specification is made by considering the data to be longitudinal, instead of cross sectional. In this case, each block group is the unit and month-year is time specification. The treatment in this model is a continuous variable. Any time unit prior to the introduction of first Airbnb is marked zero, and varies for every block group. From thereon, the value of this variable is updated by one, starting from the time point at which each subsequent Airbnb is introduced. This would allow us to study whether the effect of Airbnbs on crime is cumulative over time.

5 Results

5a Visualizations

We begin by looking at visualizations of crime data. The plots in Figure 9 plot crime frequencies pre- and post-treatment, for each crime type, as well as total crime. We see that the slopes of the blue and orange lines are not observably different in any plot. Crossing of the lines, which indicates the time point at which the two slopes interchange, is not closely following the treatment in any of them either. Time point 4 indicates discontinuity. For most types, the lines intersect before treatment time. For Drunkenness, Robbery and Vandalism, post-treatment line is tapering through the time period 4-7, indicating a general decrease in crime rate. It crosses over after the introduction of Airbnb. However, the

As an alternative, we also considered boxplots of the same data. However, for half of the crime types, the entire box is exactly at zero. This does not give us any information on the change in crime trend post treatment.

5b Exploratory analyses

Next we investigate this data through quantitative analysis to further study the pattern using paired t-tests with one-sided hypothesis. For this purpose, we remove crime data for the treatment month, and pair remaining observations for each block group at (t-k, t+k) time period. For example, crime data for three months before and three months after the introduction are considered to be correlated or pre-post observations on the same unit. An important concern here is that of autocorrelation causing dependence among observations. However, for the purpose of preliminary investigation, we set aside this limitation. Most of the crimes have maximum frequency at 1-2 occurrences per month, also raising concerns about the normality assumption. The results are presented below nonetheless.

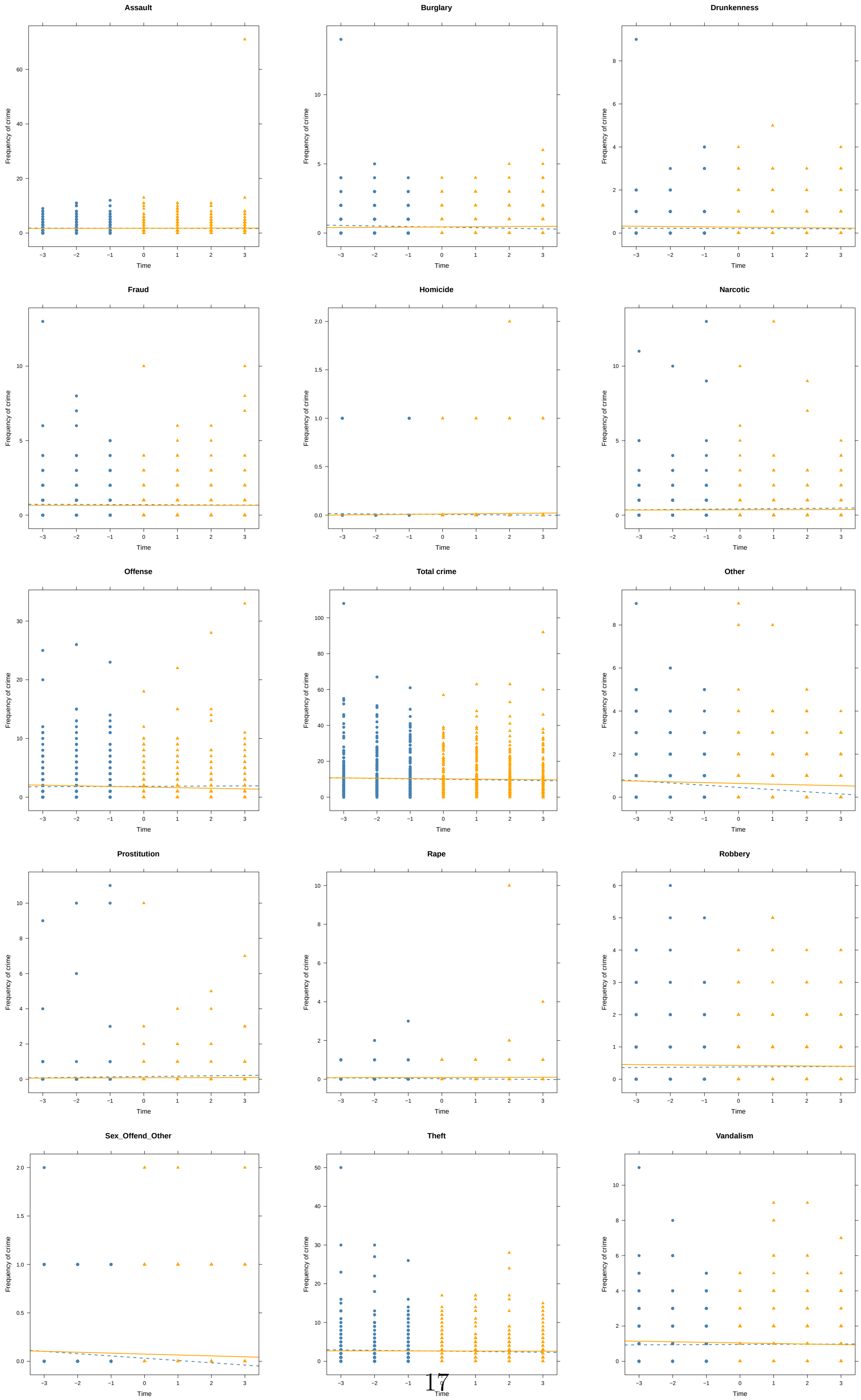


Figure 7: Crimes pre and post treatment

Crime	$H_1 : \mu_{t^-} > \mu_{t^+}$	$H_1 : \mu_{t^-} < \mu_{t^+}$
Assault	0.41	0.59
Burglary	0.82	0.18
Drunkenness	0.28	0.72
Fraud	0.83	0.17
Homicide	0.25	0.75
Narcotic	0.59	0.41
Offense	0.98	0.02
Other	0.98	0.02
Prostitution	0.81	0.19
Rape	0.03	0.97
Robbery	0.22	0.78
Sex offense	0.92	0.08
Theft	0.92	0.08
Vandalism	0.29	0.77
Total	0.96	0.04

Table 1: p-values from t-tests

In table 1 we have highlighted all p-values less than 0.1. Interestingly, more tests with the 'greater than' type alternative hypothesis are significant, indicating that the discontinuity caused due to introduction of Airbnb, in fact, increased crime of certain types (Offense, Other, Sex offense (other than prostitution), Theft and Total crime) in block groups.

However, these tests can be misleading due to the relatively short pre- and post-treatment time periods, as well as the pooling of block-groups. We show results from an alternative specification below in Figure 8. Each panel in this figure presents the distribution of estimated differences in post- vs. pre-treatment means for a particular crime across block-groups. In other words, each data point underlying a distribution is the estimated difference in means pre- vs. post-treatment for a single crime type with regards to a single block-group. The average estimated difference in means is represented by a solid gray line, while the observed 5^{th} to 95^{th} and 2.5 to 97.5 quantiles are shown as wide and narrow dashed red lines.

The pre-treatment period for each block-group X crime combination is 6 months long, while the accompanying post-treatment period is the treatment month and 6 months thereafter. The t-tests are not paired as only the pairing of the 6 months pre and 6 months post values makes sense. We can see that for all crime types the empirical 90 and 95% CIs include 0.[1] Further, for many crime types the mean estimate is indistinguishable from zero and every distribution is sharply peaked around that point.

[1]Prostitution may appear to not include zero, but in fact does. 11 block-group estimates were negative, 6 were positive, and the remaining 140 were exactly zero. This is an artifact of the limited number of reports of prostitution across the city.

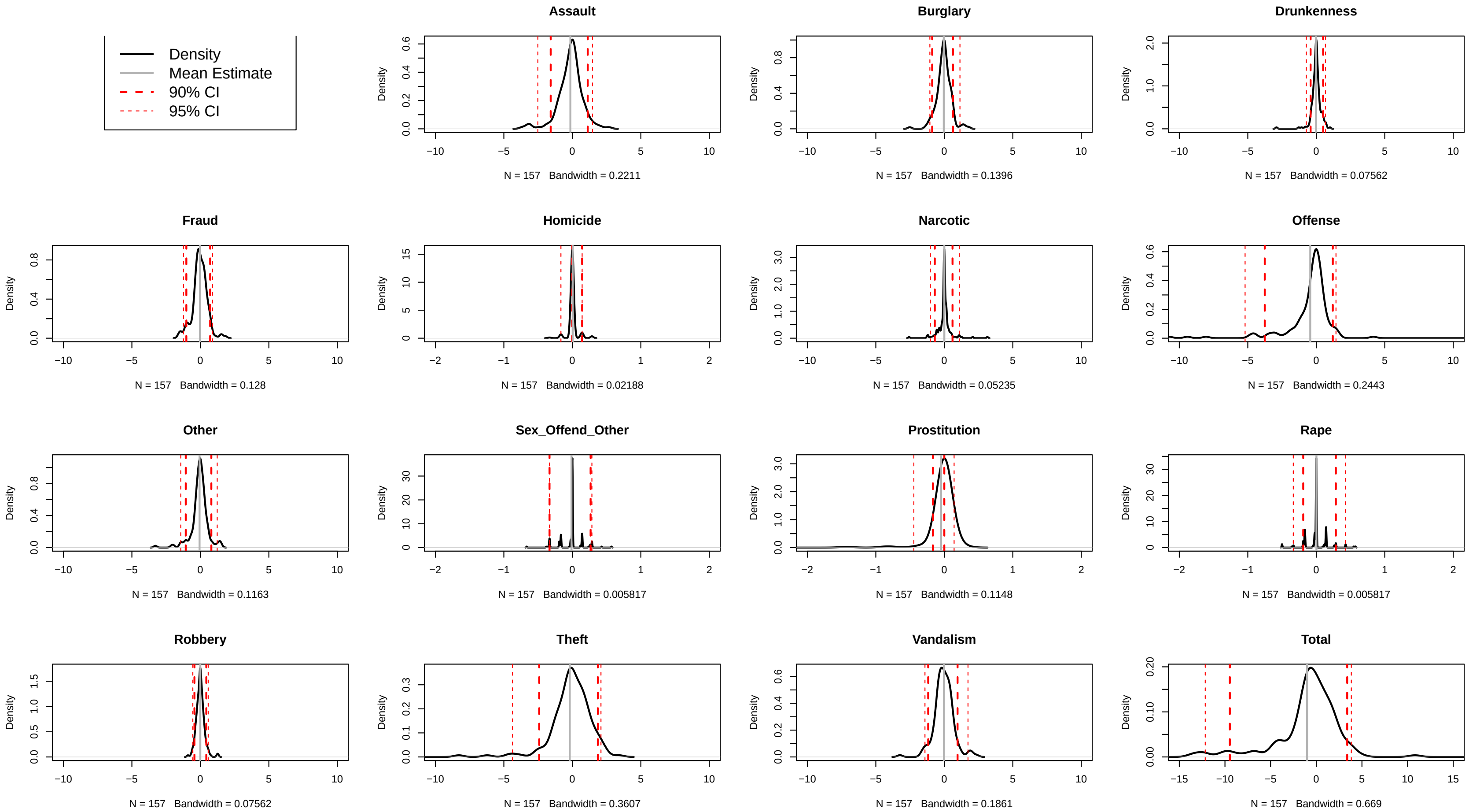


Figure 8: Block-Group x Crime t-test Difference in Means Estimates (12mo window)

5c Models

We begin by looking at the p-values of coefficients estimated using standard multiple linear regression, in which outcome is assumed to be a Gaussian variable. Values significant at 10% are marked in Red for ease of read.

Crime	Time Coefficient	Treatment Coefficient
Assault	-0.01	-0.03
Burglary	0	-0.04
Drunkenness	-0.01	0.07
Fraud	0	-0.05
Homicide	0	0
Narcotic	0.01	-0.05
Offense	-0.07	0
Other	$-0.05^{\cdot}$	0.12
Prostitution	0.01	-0.05
Rape	0	0.05
Robbery	0	0.06
Sex offense	$-0.01^{\cdot}$	0.03
Theft	-0.04	-0.06
Vandalism	-0.02	0.13
Total	-0.19	0.18

Note: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 2: Estimates from linear regression

Table 2 indicates that using linear-model-based regression discontinuity analysis, none of the crime types seems to have changed in frequency of occurrence, as a result of introduction of Airbnbs in block groups. The R-squares of these regressions are very low (< 1%) indicating that the results may not be reliable. Clearly, there are other factors at play affecting crime. If we change the distribution of outcome variable to be Poisson, the qualitative results remain the same. In these generalized linear models, the coefficients for time variable for crime categories 'Other' and 'Other Sex Offense' are significant at 10%. For total crime, the same is significant at 5%.

In the next set, we see coefficients from models with block group fixed effects. The r-square of MLR with fixed effects is significantly higher. For different crime types, it ranges between 0.2 and 0.7, with models for Prostitution and Theft taking the highest values (¿ 0.7). The model R-square is 0.81 for total crime. However, some of these models have difference of close to 20% in the r-square and the adjusted r-square, indicating limited out-of-sample validity. This phenomena usually indicates that too many coefficients are being estimated for the given sample size. This, of course, reduces the overall reliability of the model. Interestingly, none of the treatment effects are significant for any of these models. The time variable is once again significant for crime categories 'Other' and 'Other Sex Offense' at 5% and 10% respectively. A large number of block group fixed effects are significant in each of these models, indicating that there are block group-level characteristics aside from Airbnbs that affect crime in the neighborhood.

When we change the distribution of crime to Poisson, fewer block group fixed effects are significant, and the R-square value is lower. However, some of the main effects are significant and tabulated below. Once again, values significant at $p < .1$ marked in red.

Crime	Time Coefficient	Treatment Coefficient
Assault	0.01	-0.02
Burglary	0	-0.09
Drunkenness	-0.05	$0.3^{\cdot}$
Fraud	0	-0.08
Homicide	0.1	-0.11
Narcotic	0.03	-0.13
Offense	-0.04	0.01
Other	-0.09^{**}	0.2
Prostitution	0.11^{***}	-0.57^{***}
Rape	-0.02	0.65^{**}
Robbery	-0.01	0.16
Sex offense	-0.21^{***}	$0.45^{\cdot}$
Theft	-0.01	-0.02
Vandalism	-0.02	0.13
Total	$-0.02^{\cdot}$	0.02

Note: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 3: Estimates from Poisson regression with block group fixed effects

We see that prostitution is the only crime type that has significant main effects for time as well as treatment. It indicates that the overall, prostitution increases from 3 months before treatment to 3 months after. However, on an average, prostitution drops significantly after the introduction of Airbnb. This model specification indicates that treatment significantly changes three other types of crime. Drunkenness, Rape, and other Sex Offenses significantly increase after introduction of Airbnbs in block groups. The pattern here is slightly discerning, yet quite evident. All types of sex offenses show an upward trend in either time or treatment status. But we cannot be too sure since the signs are exactly opposite for the two coefficients. The fitted values vs. residuals plots for all models indicate that the residuals are heteroskedastic and positively skewed. Although this model specification provides some interesting results, the fit can be improved by making further modifications.

5d Validation

5d.1 Robustness

We fit all of the models specified above to two other data sets to check for robustness. Instead of looking at three months prior to and post treatment, we consider crime for 1 month (681 unique block groups) and 6 months (157 unique block-groups). The general findings persist in the simple linear regressions. As for the generalized linear model with poisson family and block fixed effects, some treatment effects are significant. For the one month data, crime types Drunkenness, Homicide, Other, and Robbery are significantly affected by the introduction of Airbnb. Whereas if we compare the average crime for six months before and after the treatment, Prostitution, Rape, and Total crime are significant. Given the many issues that surround the one month data, and the fact that results from six month data mirror the original dataset, we consider the results robust. Additionally, the pattern of fitted values vs. residual plots is very similar across the two robustness analysis.

5d.2 Multiple testing

Separating out crime types and fitting separate models on each of them allows us to look at the potential impact of introduction of Airbnb on certain crimes in the neighborhood. However, these can be considered to be multiple realizations of the effect of Airbnb on crime in Philadelphia. Therefore, we can use multiple testing assuming the null hypothesis 'introduction of AirBnbs does not affect crime in the neighborhood'. The literature does not provide consistent guidance on when to use multiple testing procedure. [Rice, 1989] discusses this issue in great detail. Our decision to use the Bonferroni test across models for different crime types relies on the second category of tests that the author considers it is appropriate to apply simultaneous testing; where separate tests are conducted to address a common null. The tests for each crime type cannot be pooled without loss of information. However, main hypothesis can only be rejected if at least one of the component tests is significant. Therefore we employ the Bonferroni adjustment.

It is held by some researchers that 'Bonferroni adjustments are, at best, unnecessary and, at worst, deleterious to sound statistical inference' [Perneger, 1998]. Inference based on this method relies on the number of multiple tests conducted. However, we are using it as a tool to determine the validity of results discussed in the earlier subsection. Additionally, given that the treatment effect was significant in a small number of models, this adjustment would not worsen the probability of type I error. At most, it will affect the probability of type II error negatively. However, that discussion is beyond the scope of this paper.

When testing at $\alpha = 0.05$, for 15 tests, P(at least one false positive) is almost 0.55, for 20 tests it is over 0.6 [Cabin and Mitchell, 2000]. Therefore it is useful to use Bonferroni adjustment in this analysis. The adjustment works so that an individual test (out of a total of k tests) is considered significant at level $1 - \alpha$, if the p-value is lower than α/k. We have 14 distinct crime types, and we are interested in significance at 10%. Therefore p-value below 0.007142 is considered significant. If we look at the final model (poisson outcome with block fixed effects), Rape and Prostitution are the only two crimes which were significantly different from zero post-treatment, over and above the baseline, which is the pre-treatment period.

5d.3 Multivariate multiple regression

As an experiment, we try to fit a multivariate multiple regression on this data. There is possible dependence across crime types, which deserves investigation. Such dependence can arise due to common causes mentioned in the explanation of block group fixed effects, or due to common criminals, geographical locations being more prone to certain types of crimes than other, tourism patterns, or other socioeconomic factors. Given this, it is valuable to consider different types of crimes as a matrix of dependent variables. Such model also accounts for covariance across these outcome variables.

For testing purpose, we provide a matrix of counts of 14 crime types as the dependent variable. Independent variables are time and treatment indicator. We then conduct a Multivariate Analysis of Variance (MANOVA) on the coefficients of the fit. The null hypothesis is 'none of the crime models differ significantly based on treatment or time coefficients'. Using the Wilks' statistic, we do not see significant effect, indicating that even when modeled as multiple correlated variables, the introduction of Airbnbs does not affect crime in the neighborhood. This corroborates with our prior findings.

5e Panel data analysis

Finally, we analyze this data as a panel. We specify a panel data model using the plm package in R, while exploring Individual only, Time only, and Twoways fixed effects. In addition to the effect of a continuous treatment, defined by the introduction of Airbnbs in a block group, we also estimate a parameter for lagged crime (lagged by one month). We tabulate results for three important crime types, two of which have consistently shown significant effect (Prostitution, Rape), and total crime.

Crime	Treatment (rank ordered)	Lagged crime	R-square
Prostitution (Individual)	-0.01*	-0.03**	0
Prostitution (Time)	0	-0.39***	0.15
Prostitution (Twoways)	0	0.03**	0
Rape (Individual)	0	-0.04***	0
Rape (Time)	0	0.03***	0
Rape (Twoways)	0	-0.04***	0
Total (Individual)	-0.4***	0.1***	0.02
Total (Time)	0.1.	0.8***	0.64
Total (Twoways)	-0.15***	0.07***	0.01

Note: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 4: Estimates from Panel data analysis with three fixed effects specifications

We notice that only the models with time fixed effects i.e. fixed effects for each year-month combination have non-negligible explanatory power, for the coefficients to be reliable. Of these, only total crime is significantly affected (at 10%) by introduction of Airbnbs. This pattern of results is maintained across crime types.

These results contradict an earlier understanding that controlling for block group fixed effects improves estimation as well as model performance.[2] However, the residuals from panel models are not well-behaved either. Additionally, it looks like the lagged crime frequency is always contributing to the explanatory power of the model.

[2]It should be noted that the model fit using two-way fixed effects vastly improves when the dependent variable is differenced. The magnitude and significance of the main effects remain largely unchanged though.

Therefore it is not appropriate to exactly compare the two models.

6 Limitations

We have consistently failed to reject the null hypothesis of no effect on crime as a result of Airbnb introduction. These findings are robust across crime types, time-periods, and methodologies. Further, findings contradicting the null, often disappear when the time period is extended: e.g., prostitution decreases significantly following the introduction of Airbnb using a pre-/post-treatment window of ± 3 months, but this finding vanishes in the ± 6 month block-group X crime t-tests and all forms of the panel analyses. While robust to alternative specifications, our findings certainly do not provide a definitive answer to the question of Airbnb's impact on crime. There is a sizable set of potential confounders that we were not able to address.

6a Unobserved Treatment and Survivorship Bias

Perhaps the most problematic confounder, and also the most difficult to adjust for, relates to fact that we cannot be sure that we have observed the true treatment. We follow Zervas et al. [2016] in estimating Airbnb "availability" (i.e., introduction) based on the submission of a review. However, this is not completely reliable for a number of reasons. Firstly, as noted by Zervas et al. [2016], roughly two-thirds of Airbnb stays result in a review and reviews can be submitted up to 14 days after occupancy. This means that there is a high probability that we *observe* treatment as occurring after an Airbnb unit has already been occupied by guests.

Another problem, which Zervas et al. [2016] fail to mention, is that when the Airbnb site is scraped, deleted units remain unobserved. Thus it is impossible to determine if a city/block-group was treated at the time of the first *observable* review, or if treatment truly occurred with the occupancy of another listing that was then deleted prior to a researcher's data collection. The further away data collection occurs from the true time of treatment, the higher the probability of observing a secondary treatment unit as the initial treatment becomes.

This is further compounded by the introduction of potential for survivorship bias, whereby units that were deleted are fundamentally different from those that remained open. It is easy to imagine that Airbnb units with low ratings are far more likely to be removed from the site by their owners (and subsequently re-listed under a new ID) than highly rated units, which would subsequently feed into occupancy rates and, in theory, crime rates. To demonstrate the reality of these problems, Figure 9 shows the frequency of listing deletion by date of listing creation, for units in Philadelphia proper that were deleted between September of 2016 and December of 2016. These comprise 93 of the 2,682 ($\sim$ 3.5%) units within Philadelphia proper originally collected in September, 2016.

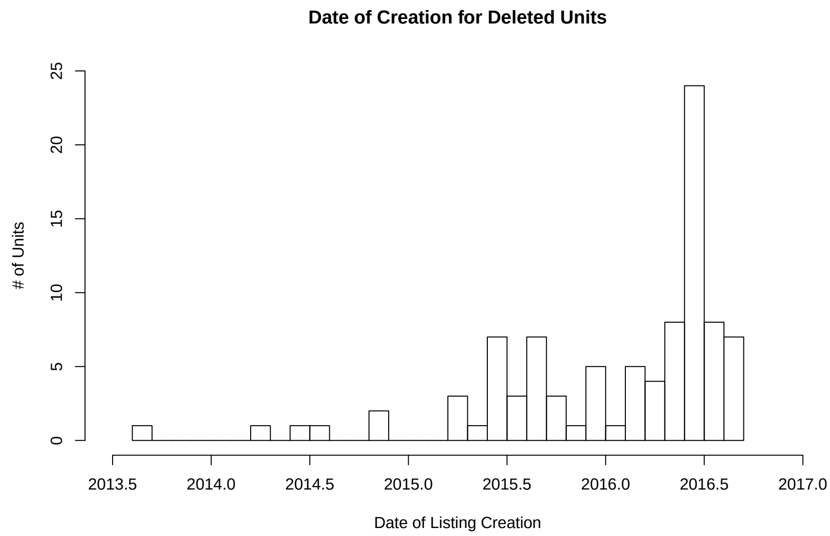


Figure 9: Creation Dates for Listings Deleted Between Sept. 2016 and Dec. 2016

Over just a three month span, listings from as far back as September, 2013 were removed from the site, as well as seven listings from September, 2016 – when we collected our data. It is not unreasonable then to think that over a longer span of time the deletion of listings could have a serious impact on estimates associated with the introduction or existence of Airbnb units.

In terms of the volume, of all listings in the Philadelphia metropolitan area, 494 units were shut down over this time period. However, of the total 910 units which belonged to a valid block group and had crime data available, only 84 were closed. Finally, as described in the paper, we further adjust the data so that only those units which were 'first Airbnb unit introduced in a block group' are used for statistical analysis. From this pool of 191 units, we lose only 21 units. These numbers are significant from an overall effect point of view. If close to 10% units drop out of the sample, it is possible that the effect on crime may look different. However, for the analysis in this paper, we do not address this concern. Given the popularity of Airbnb and our study of a large metropolitan such as Philadelphia, we assume that another Airbnb would be introduced in the block group within a short period of time, not affecting our treatment.

6b Reverse Causality

Another limitation of this study is that there is non-zero probability of reverse-causality. Imagine for a moment that you are considering opening up an Airbnb unit in your neighborhood. If your neighborhood is relatively safe then perhaps you are more likely to operate an Airbnb. However, if your neighborhood has a relatively high crime rate, you may be more reluctant to try out hosting on Airbnb. As the potential Airbnb host in the high crime neighborhood, you start to see crime rates decreasing in your area and thus decide to open an Airbnb, which (dependent on timing) may be observationally equivalent to your opening resulting in a reduced crime rate.

6c Omitted Variable Bias

Finally, there is the potential for our estimates to be the result of omitted variable bias. We have not considered additional contextual data about the block groups (e.g., income, age distribution, racial composition, community-police relations, availability of transportation). Other traits of the units-of-analysis present a myriad of potential confounding covariates that would need to be controlled for given the non-random assignment of Airbnb units.

7 Conclusion

Based on multiple analyses using several different specifications of the treatment phenomenon as well as the outcome, we conclude that the introduction of Airbnbs does not affect crime in a pre-specified neighborhood of the unit. Although our analysis did not detect any large scale changes in the crime rate as a result of introduction of Airbnb listings to the neighborhoods, the social changes as a result of this growing economy cannot be neglected. This research was a first step towards the assessment of a new form of landscape in an increasingly complicated urban setting.

Our research was intended to address the ongoing confusions around the possibilities of both negative and positive impacts of Airbnb listings on the social climate of different neighborhoods. While the literature on real estate and economy have shown some impacts on the economic conditions of neighborhoods, particularly in weaker communities Lee [2016], these economic changes have not shown to influence the safety of neighborhoods. One reason to this issue might be the fact that most of the Airbnb listings in our case study were introduced in 2016 which does not allow us to fully investigate the long-term effects of these listings. Our analysis showed that there are no short-term social issues associated with the emergence of Airbnb listings and this business can be practiced in neighborhoods without any significant concerns around the safety of neighborhoods. The most important short-term influence that we expected to observe in this study was the impact of increased tourist foot traffic in different neighborhoods, particularly the less known neighborhoods. As discussed in the reverse causality section, however, the possible impacts of these listings in the crime prone neighborhoods still begs further investigation.

We suggest the future work to consider investigating the Airbnb impacts where the listings can be found more evenly distributed across different neighborhoods. The future research may also consider the spatial relations through a point pattern process so as to avoid the limitations of block group level aggregation and spatial boundaries. Considering other geographic and cultural contexts is another issue that may show different results particularly those areas where lacked hospitality related commodities prior to Airbnb. Our results benefit urban sociologists, planning and landscape studies, real estate, hospitality and tourism literature, Airbnb and other similar industries.

References

Andreas M Kaplan and Michael Haenlein. Users of the world, unite! the challenges and opportunities of social media. *Business horizons*, 53(1):59–68, 2010.

The Economist. The rise of the sharing economy. `http://www.economist.com/news/leaders/21573104-internet-everything-hire-risesharing-economy`, 2013.

Rolfe Winkler and Douglas Macmillan. The secret math of airbnb?s $24 billion valuation. http://www.wsj.com/articles/the-secret-math-of-airbnbs-24-billion-valuation-1434568517, 2015.

Airbnb. https://www.airbnb.com/about/about-us, 2016a.

PlanetDen. Top 10 biggest hotel chains in the world. http://planetden.com/hotels-restaurants/top-10-biggest-hotel-chains-world, 2015.

InterContinental Hotels Group. IHG annual report. https://www.ihgplc.com/, 2015.

Airbnb. Airbnb in nyc, housing report. http://nycommunities.org/sites/default/files/Airbnb\%20in\%20NYC.pdf, 2015.

Airbnb. The economic impacts of home sharing in cities around the world. https://www.airbnb.com/economic-impact, 2016b.

Paula Bialski. *Becoming intimately mobile*. Peter Lang London and New York, NY, 2012.

Kristen Lepore Southern California Public Radio. Apartment conversions to airbnb hotels driving up la rents. http://www.scpr.org/news/2015/03/16/50321/are-apartment-conversions-to-airbnb-hotels-driving/[http://perma.cc/PF4SKBST], 2015.

George Arnett James Ball and Will Franklin. London's buy-to-let landlords look to move in on spare room website airbnb. https://www.theguardian.com/technology/2014/jun/20/buy-to-let-landlords-leasing-properties-airbnb-uk, 2014.

Dayne Lee. How airbnb short-term rentals exacerbate los angeles's affordable housing crisis: Analysis and policy recommendations. *Harv. L. & Pol'y Rev.*, 10:229, 2016.

Adrian Kudler. Meet la?s most prolific airbnb host with 78 units for rent. `http://la.curbed.com/archives/2015/03/airbnb_los_angeles_most_prolific_host_ghc.php[http://perma.cc/4WMD-7MXH]`, 2015.

Maureen Kennedy and Paul Leonard. *Dealing with neighborhood change: A primer on gentrification and policy choices.* Brookings Institution Washington, DC, 2001.

Benjamin G Edelman and Michael Luca. Digital discrimination: The case of airbnb. com. *Harvard Business School NOM Unit Working Paper*, (14-054), 2014.

Michael Todisco. Share and share alike? considering racial discrimination in the nascent room-sharing economy. *Stanford Law Review Online*, 67:121, 2015.

Anti eviction Mapping Project. The commodification of everyday life: This bed?s for sale. `http://www.antievictionmappingproject.net/airbnb.pdf`, 2014.

Steven Leigh Morris. Airbnb is infuriating the neighbors. is it time for new rules? `http://www.laweekly.com/news/airbnb-is-infuriating-the-neighbors-is-it-time-for-new-rules-5343663[http://perma.cc/4JG2-KAJM]`, 2015.

Dennis W Roncek. Dangerous places: Crime and residential environment. *Social Forces*, 60(1):74–96, 1981.

Stephanie W Greenberg and William M Rohe. Neighborhood design and crime a test of two perspectives. *Journal of the American Planning Association*, 50(1):48–61, 1984.

James Byrne. Cities, citizens, and crime: The ecological/nonecological debate reconsidered. 1986.

Robert J Sampson. Crime in cities: The effects of formal and informal social control. *Crime and justice*, pages 271–311, 1986.

Dennis W Roncek and Mitchell A Pravatiner. Additional evidence that taverns enhance nearby crime. *Sociology and Social Research*, 73(4):185–188, 1989.

Robert J Sampson and W Byron Groves. Community structure and crime: Testing social-disorganization theory. *American journal of sociology*, pages 774–802, 1989.

Charles Swartz. The spatial analysis of crime: What social scientists have learned. *Analyzing crime patterns: Frontiers of practice*, pages 33–46, 2000.

Rutherford H Platt and Oscar Newman. Defensible space: Crime prevention through urban design, 1975.

Jane Jacobs. *The Death and Life of Great American Cities*. Vintage, New York City, NY, USA, 1961.

Dennis P Rosenbaum. Community crime prevention: A review and synthesis of the literature. *Justice Quarterly*, 5(3):323–395, 1988.

Douglas D Perkins, Abraham Wandersman, Richard C Rich, and Ralph B Taylor. The physical environment of street crime: Defensible space, territoriality and incivilities. *Journal of Environmental Psychology*, 13(1):29–49, 1993.

Simon Chih-Feng Shu et al. Housing layout and crime vulnerability. *Urban Design International*, 5(3-1):177–188, 2000.

Simon Shu and Jason Huang. Spatial configuration and vulnerability of residential burglary: A case study of a city in taiwan. In *Proceedings, 4th International Space Syntax Symposium, London*, 2003.

Bill Hillier and SC Shu. Crime and urban layout: the need for evidence. *Secure foundations: Key issues in crime prevention, crime reduction and community safety*, 224, 2000.

Michael A Jones and Mamoun F Fanek. Crime in the urban environment. In *Proceedings of First International Symposium on Space Syntax*, volume 2, 1997.

Bill Hillier and Julienne Hanson. *The social logic of space*. Cambridge university press, 1989.

Wesley G Skogan, Susan M Hartnett, Jill DuBois, Jennifer T Comey, Marianne Kaiser, and Justine H Lovig. *On the beat: Police and community problem solving*. Citeseer, 1999.

Robert C Trojanowicz and Robert Baldwin. *An evaluation of the neighborhood foot patrol program in Flint, Michigan*. Michigan State University East Lansing, 1982.

James Q Wilson and George L Kelling. The police and neighborhood safety: Broken windows. *Atlantic monthly*, 127(2), 1982.

George L Kelling. Order maintenance, the quality of urban life, and police: A line of argument. *Police leadership in America*, 296, 1985.

Jack R Greene and Ralph B Taylor. Community-based policing and foot patrol: Issues of theory and evaluation. *Community policing: Rhetoric or reality*, pages 195–224, 1988.

Jack R Greene and Stephen D Mastrofski. *Community policing: Rhetoric or reality*. Praeger New York, 1988.

Denise C Gottfredson, Richard J McNeil, and Gary D Gottfredson. Social area influences on delinquency: A multilevel analysis. *Journal of Research in Crime and Delinquency*, 28(2):197–226, 1991.

Anastasia Loukaitou-Sideris, Robin Liggett, Hiroyuki Iseki, and William Thurlow. Measuring the effects of built environment on bus stop crime. *Environment and Planning B: Planning and Design*, 28(2):255–280, 2001.

Lawrence E Cohen and Marcus Felson. Social change and crime rate trends: A routine activity approach. *American sociological review*, pages 588–608, 1979.

Anthony E Bottoms and Paul Wiles. *Environmental criminology.* The Oxford handbook of criminology, 1997.

Lawrence W Sherman. Hot spots of crime and criminal careers of places. *Crime and place*, 4:35–52, 1995.

Richard L Block and Carolyn Rebecca Block. Space, place and crime: Hot spot areas and hot places of liquor-related crime. *Crime and place*, 4(2):145–184, 1995.

Michael E Buerger, Ellen G Cohn, and Anthony J Petrosino. Defining the ?hot spots of crime?: Operationalizing theoretical concepts for field research. *Crime and place*, 4:237–257, 1995.

Wilpen Gorr and Andreas Olligschlaeger. Crime hot spot forecasting: Modeling and comparative evaluation, summary. *Washington, DC: National Criminal Justice Reference Service*, 2002.

William R Rice. Analyzing tables of statistical tests. *Evolution*, 43(1):223–225, 1989.

Thomas V Perneger. What's wrong with bonferroni adjustments. *Bmj*, 316(7139): 1236–1238, 1998.

Robert J Cabin and Randall J Mitchell. To bonferroni or not to bonferroni: when and how are the questions. *Bulletin of the Ecological Society of America*, 81(3): 246–248, 2000.

Georgios Zervas, Davide Proserpio, and John Byers. The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry. *Boston U. School of Management Research Paper*, (2013-16), 2016.