Spatio-temporal and socio-economic analysis of Chicago parking tickets

Jin Guo Ye Chao Yilong Liu

{jnguo yechao yilliu} @ ucdavis.edu

1 Introduction

Based on the data from Pro Publica, in 2016, the Chicago government brought in 264 million of dollars revenue by issuing parking ticket, which is 7% of the total operation cost of the city (Sanchez & Kambhampati, 2018). In 2017, 162 million dollars of tickets more were issued, and 87 million dollars late fees were charged (Black, 2018). According to Top View, average 97.2 dollars ticket revenue were made by the government from each capita in Chicago for each year which is the 4th highest among the United States (Thomas, 2018).

Managing vehicles and parking system by issuing ticket is understandable but the Chicago government used it as one way to make profit and cause some serious debt problems. To fill the budget gap, the government decided to raise sticker violation fine to 200 dollars in 2012. This ticket became the largest source of the debt in the following years due to its high cost and large quantity (Black, 2018). In 2015, Chicago's unpaid ticket grew to 1.5 billion dollars which was higher than the ticket debts in Los Angeles and New York even though these two cities had higher population and more ticket issued annually (Brockway, 2015). Nearly 40% of the total debt was from low-income families who couldn't afford it (Sanchez & Kambhampati, 2018). Moreover, the government decided to punish the people who did not pay the ticket by suspending their driver license and impound their vehicle. Also, the people with unpaid tickets were prohibited to find a job and get any contract, license, and social welfare until they clear their debts. (Sanchez & Kambhampati, 2018).

Not only the ticket debt has become one of the biggest burdens for the citizens in Chicago, but also the government has been frustrated with these ticket debts for a long time. Even the government used several methods to punish the people with huge ticket debt and forced them to clear their debts. These people also had some tactics to help them get away from the punishments. One of the most often used method in Chicago was Applying Chapter 13 Bankruptcy. By applying this type of bankruptcy with only 310 dollars, debtors could free their impounded cars, reinstate their suspended license, stop any debt collection activities, and wipe out some parking tickets (Sanchez & Kambhampati, 2018). Even though applying bankruptcy couldn't waive all the debt they have, debtors were allowed to pay the debt within a longer period (up to 5 years). However, based on the studies in 2010, less than a quarter of the Chapter 13 bankruptcies that included ticket debt end successfully because most debtors just tried to get their impounded cars, reinstate their licenses, or keep their jobs but not intended to clear the debts (Sanchez & Kambhampati, 2018). Although the government could go after those debtors again after the previous bankruptcies dismissed, those debtors might apply the bankruptcy again which means that cities could only recover a little from these cases. Based on the data from 2007 to 2017, number of Chapter 13 bankruptcies raised from 1000 cases to more than 10000 cases and the debt for each case raised from 1000 dollars to 3900 dollars (Sanchez & Kambhampati, 2018). It was a trend that people with huge debt tended to file the bankruptcy to avoid the punishment of not paying the parking ticket and it became a huge obstacle for the government to collect the parking ticket debt.

Thus, the motivations of this study are to relieve some of the debt burden for Chicago citizens and to help the government stem the flood of bankruptcy and retrieve the parking ticket debts. In this study, the spatial distribution of the parking tickets and the relationship between number of issued parking tickets and issue time are investigated which can be used to reflect the parking condition for different places and at different time. Future studies can use this information to optimize the parking system in Chicago. Secondly, the most correlated factors to number of parking tickets issued were determined in this study which can be used to design a more reasonable fine system. Lastly, the analysis on the number of bankruptcies annually in 2016 and 2017 is performed to understand the trend change.

2 Methodology

2.1 Datasets

The datasets used in this study were Chicago parking violation information, Chicago social econometrics in 2016, and Chicago zip code geographical coordinates. Details were as follow:

2.1.1 Chicago Parking Violation

Chicago parking violation information was recorded from 2016-01-01 to 2018-05-14, including violation location, ticket fine amount, violation description, license plate, and vehicle make, registration zip code, ticket status and etc. Total 2254275, 2190763, 769219 records were captured in 2016, 2017 and 2018 respectively. These records was cleaned and analyzed based on objectives in the project.

2.1.2 Chicago Social Econometrics Data

To determine correlation between violation vehicles' registration zip code and the regional attributes with over 544 variables, including household median income, population, gender distribution and average car ownership data, etc, the 2016 Chicago social econometrics dataset which primary key was zip code was brought in from TransCAD 8.0, Caliper INC, licensed by UC Davis

2.1.3 Chicago Zip Code

To visualize data in ggmap, the regional boundaries for each zip code need to be defined. Thus, Geographical polygons was brought in from Chicago City Data Portal to provide zip code regions' boundaries.

2.2 Data Pre-process

2.2.1 Violation Time Pre-process

Firstly, violation time was taken as a primary key. Then, factor type was converted into POSIXlt. After sorting the violation time chronologically, we remove the duplicating rows and subset the whole dataset by the year of 2016, 2017, and 2018 as Data_2016, Data_2017, and Data_2018 respectively.

2.2.2 Violation Location Pre-process

In the original dataset, the violation location was the address that the violations happened. For the purpose of analyzing geographical information on parking violation, these addresses were converted into longitudinal and latitudinal coordinates by using function geocode provided by google map Api. The input of the function geocode was character type address. However, inputting address in the dataset directly could cause searching error since the address names could be duplicated. After characterized the address variable in the original dataset, ", Chicago, IL" is added in the end of each address.

2.2.3 Violation Type Pre-process

The violation type variables in the original dataset lacked standardization with more than 100 levels in which there are problems with singular-plural and upper-lower case. The singular and plural was unified by using topper function and singularize function in pacman package.

2.2.4 Zip Code Pre-process

Some of zip codes are using 5-digit zip code while some of them are 9 digit (ZIP+4) format in original dataset. To match with the zip code polygon data brought in, first 5 digits of the zip code were extracted to get a unified 5-digit zip code.

2.2.5 License Plate Number Pre-process

In the original dataset, the license plate number were encrypted by algorithm. However, there are some license plate numbers didn't match with the vehicles' make in a few cases, which means one license plate number(coded) represented multiple. Also, some license plate showed up more than 360000 times. This part of data was removed so that the one license plate numbers could only match one vehicle.

2.3 Methods

2.3.1 Indexing Car Owners with Plate Numbers & Multiple Violation Record

To conduct analysis on each car owner, unique index for each car owner was created based on unique license plate numbers after modified the original dataset by using the method mentioned in 2.3.5. It means each index only corresponded to one license plate number and represent only one car owner. This method was applied on original data subsets for 2016, 2017, and 2018.

The reason is that the major subjects for this study were the people with multiple violation records. These people could be more relevant to the parking ticket debts problem. The new subsets with new index were made for 2016, 2017, and 2018 to only focus on these people who have multiple violation-record for further analysis.

2.3.2 Chicago Parking Violation Temporal Pattern Mining

The cleaned subsets of 2016, 2017, and 2018 was used to deliver the following analysis:

1. Aggregated by year For each subset, total and average tickets number and fine amount were calculated to evaluate Chicago city parking tickets issuing variation in each year.

2. Aggregated by month In each subset, monthly ticket counts andticket amount were calculated by using function tapply to evaluate Chicago city parking tickets issuing variation for each month.

3. Aggregated by day The tickets issued for each day were plotted on a calendar heat map for 2016, 2017, 2018 respectively. The calendar heat map of each year was obtained by an independent function which separates months and days in a calendar-arrange way within a whole year.

2.3.3 Chicago Parking Violation Spatial Pattern Mining

1. geocode The function geocode was offered within the google map Api by using the character type of address as an input. However, when used geocode function, the calculation speed was limited by network bandwidth and the google map server. The maximum speed only calculated 50 times per second and the limitation on the times running the function within one account is 100000. As a result, in subset of 2016, 2017, and 2018, random violation samples with after-modified address(introduced in 2.3.2) are selected to run the geocode function for a total 200000 times. The output is a dataset with information about the input address, and select the longitude and latitude information to combine with the original selected samples correspondingly to get the locations coordinates for some of the violations.

2. ggmap and Spatial Distribute After the loading ggmap with Api key, geographical coordinates could be plot onto the map. Using geom_point to get a scatter distribution plot and stat_density2d to get a distribution plot on density. In addition, by bring in the zip code boundaries geographical coordinates data (2.1.3) and merging with the yearly subsets by zip codes separately, the total tickets counts with respect to each zip code area could be visualize within the map.

3. Multi-Variable Regression on Ticket Counts per Capita In the purpose of investigating the relation between total ticket counts and socio-economic variables. The dataset of the Chicago Social Econometrics Data (2.1.2) with zip code as primary key is brought in. By

looking parking as one of the urban social demand of static traffic way and using the experience of travel demand regression, we select population, household median income, family mean income, male/female population, median age, percentage of poverty, car ownership per household and etc. as independent variables to run the multi-variable regression with dependent variable being ticket counts per capita in each zip code area. In the model section sector, according to the Akaike Information Criterion (AIC), which describes the regression quality of the statistical models. The lower the value is, the better it fits. The AIC value of the model is the following

$$AIC = 2k - 2\ln(L)$$

By testing the AIC value with different variables' combinations and running the iterations backward, we could get the function in which the AIC value is lowest and that function being the best fit to the regression. Then the most significant variables will be selected, the validity of the selected model will be checked in the end.

2.3.4 Bankruptcy Cases Recording

The bankruptcy records are selected by using sql package. Since we want to focus on the people that went into bankruptcy, in each person index, select people from bankruptcy status to non- bankruptcy and non-bankruptcy status to bankruptcy status and apply it to 2016, 2017 and 2018 subsets with unique personal index.

3 Results

3.1 Temporal Analysis

3.1.1 Analysis by year

Figure 1 and figure 2 show the aggregate counts of all the parking tickets and fine amounts in 2016 and 2017 respectively. The total ticket counts decreased by 2.8 % in 2017 compared with that of 2016. Simultaneously, the total fine amounts dropped from 257.495 million to 248.275 dollars, a decrease of 3.6%. The average fine amounts of tickets are roughly the same in 2016 and 2017, however, the statistical data until April in 2018 indicates a 28.6% drop from 2017 count, which is possibly attributed to partial exemption from late fee of the next months.
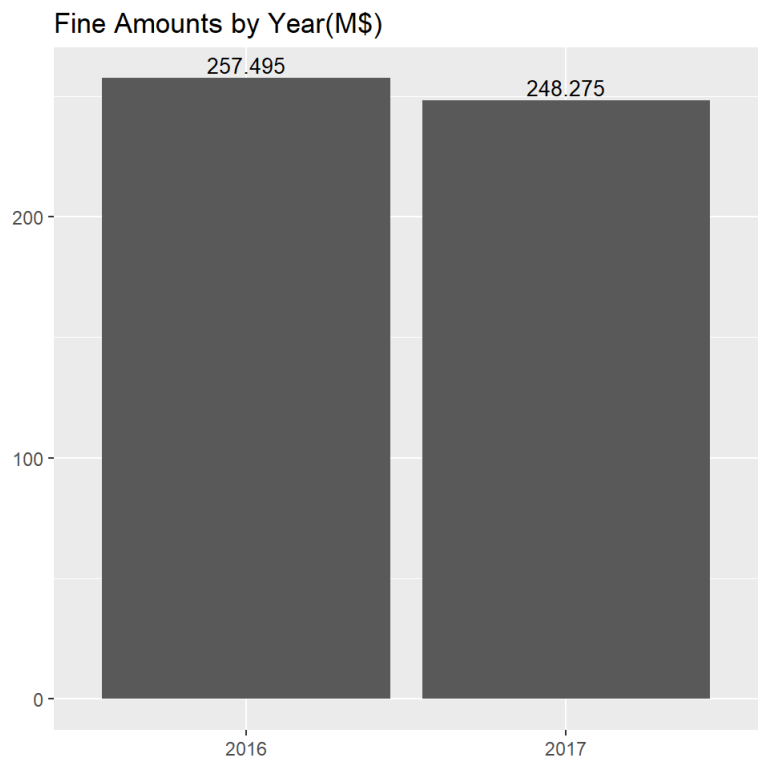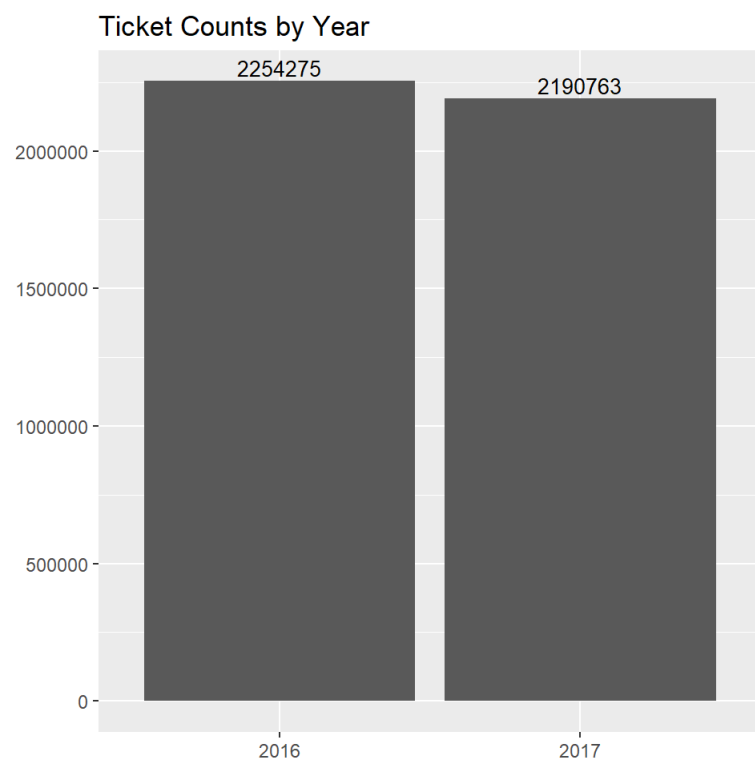

Figure 1. Annual Ticket Counts of 2016 and 2017

Figure 2. Annual Fine Amounts of 2016 and 2017

3.1.2 Analysis by month

While the aggregate counts of parking tickets and fine amounts indicates homogeneity, there exits a variance in quantity from January to December, February has the smallest number on these counts. 2016 and 2017 share the same time-varying trend both in ticket counts and fine amounts, which is a sharp decrease from January to February, then an oscillation between two peaks on March and October, finally a continuous decline on the last two months. Notably, winter months have the low rates of ticket counts and fine amounts, which is high possibly related to the cold winter weather in Chicago. There will be less trip generation in extreme weather conditions, parking, part of Static Traffic will reduce correspondingly.
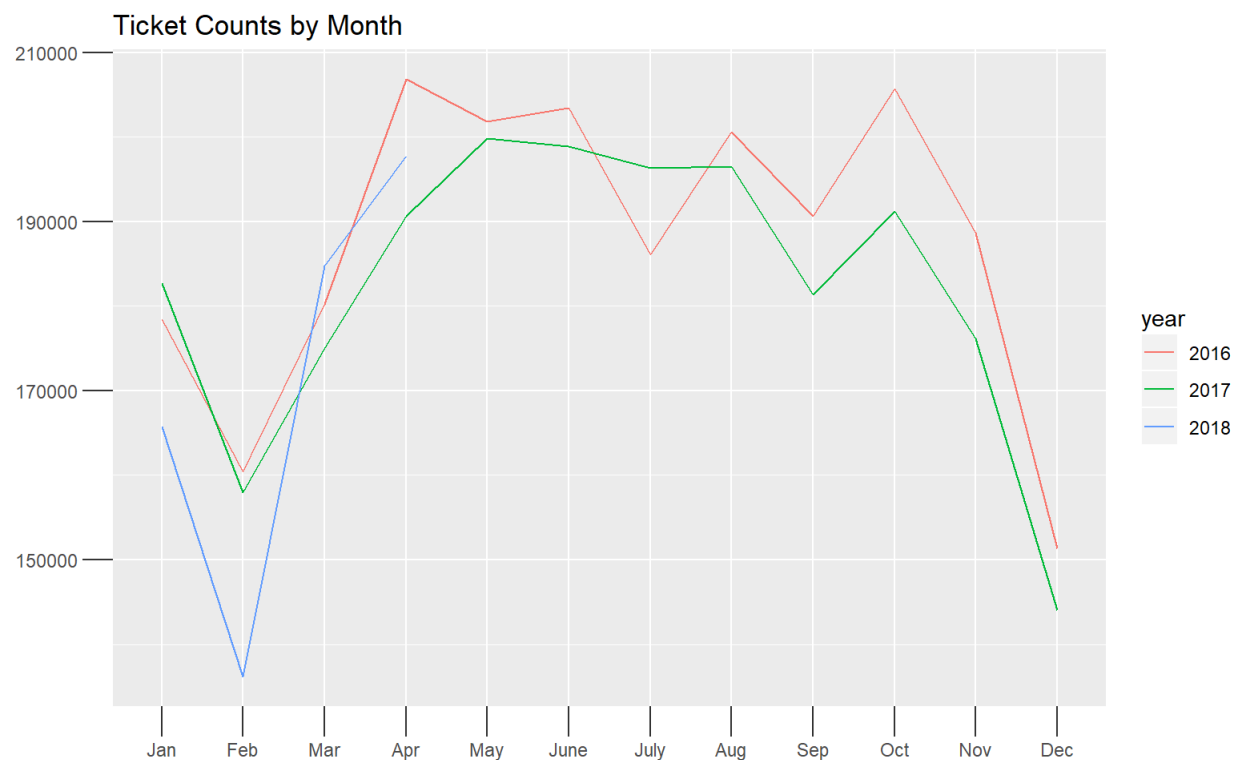
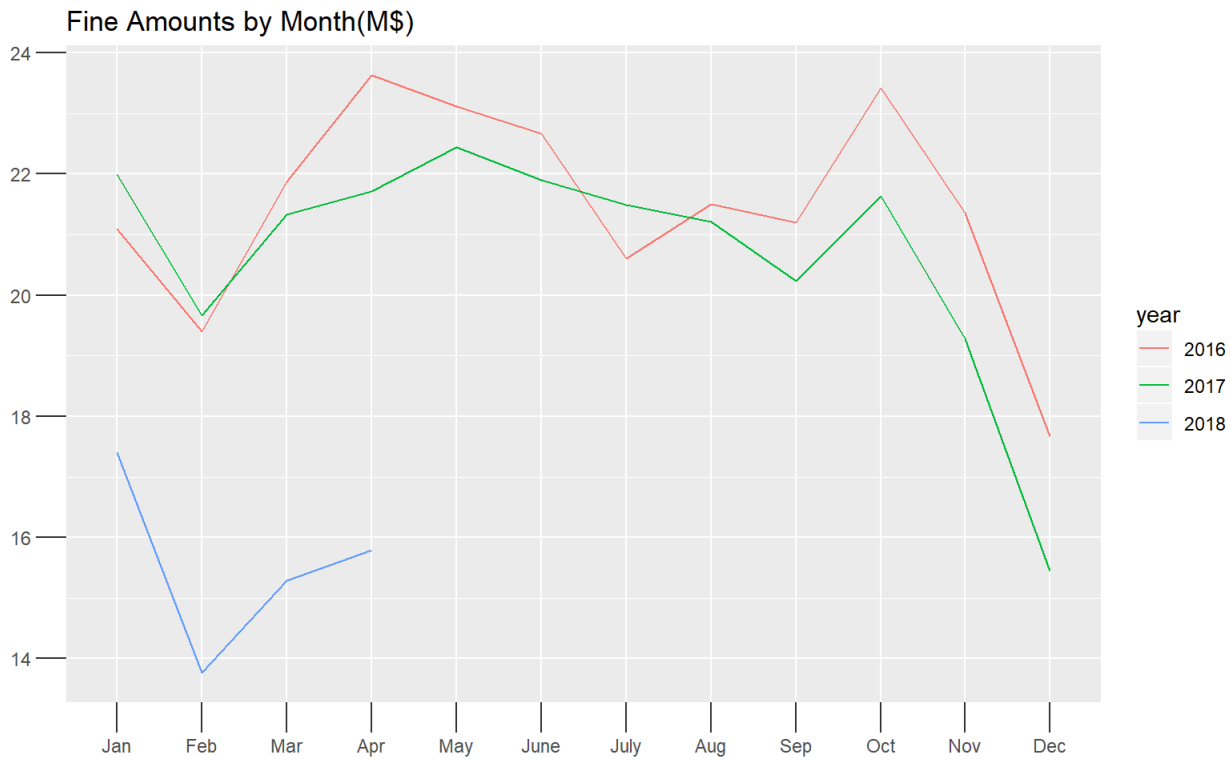

Figure 3. Ticket Counts of Each Month

Figure 4. Fine Amounts of Each Month

3.1.3 Analysis by day

The calendar heatmap shows the temporal distribution of ticket counts of each day in each year. It is obvious that weekends' parking tickets are nearly 75% less than weekdays', which suggests the main part of parking tickets are brought by commuting trips of people with work. This conclusion can be confirmed by figure 4. With the latitude and longitude information of 66000 samples of 2018 obtained by the geocode() method, the heatmap of parking tickets demonstrates the spatial heterogeneity between weekdays and weekends. The regions that always exists on the heatmap(the upper part on weekdays) are mainly residential areas, producing continuous parking needs. The lower part on weekdays becomes more sparse, even disappears on weekends. We can figure out this area is downtown Chicago, with CBD located in the center.
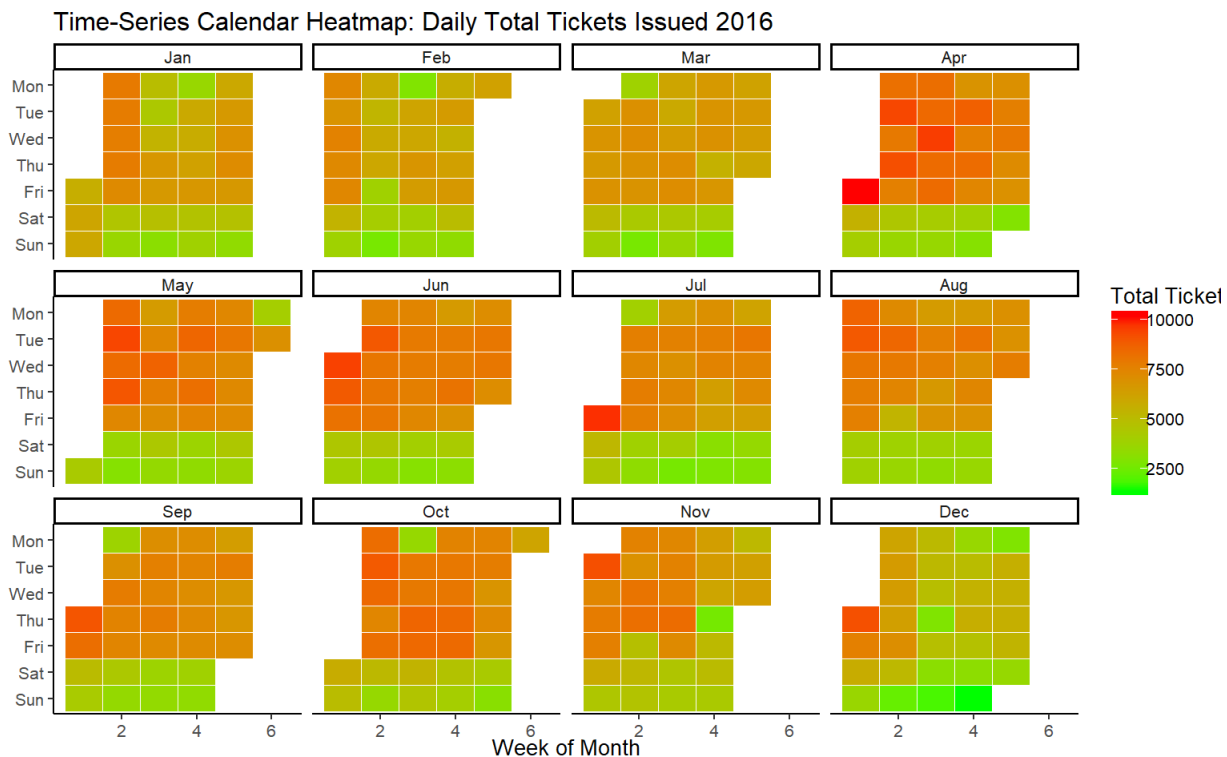

Figure 5. Calendar Heatmap of Ticket Counts in 2016

3.2 Spatial Analysis

3.2.1 Zonal distribution of parking tickets

Based on the longitude and latitude information of the 2018 samples, the spatial distribution parking tickets shows an inhomogeneity in
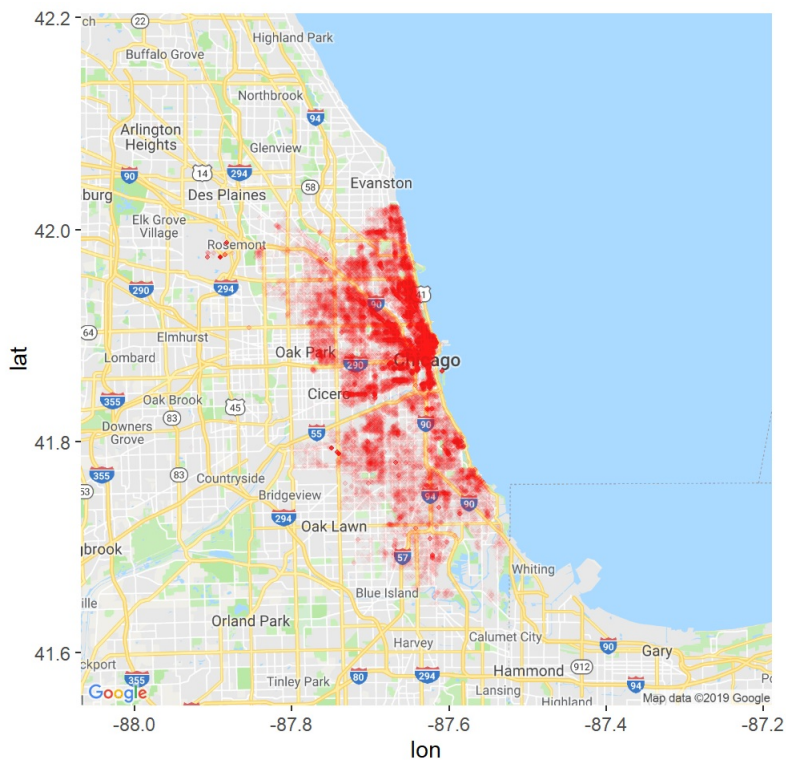
Figure 6.

Figure 6. Spatial Distriution of Ticket Counts in 2018

In order to figure out the relationship between tickets' spatial distribution and socio-economic factors, similarly to urban transportation modeling, we apply the Traffic Analysis Zone(TAZ) method to define research interest areas. Naturally, areas divided by zip codes are set as TAZs(Figure 7), for the census data we use is archived on zip codes.
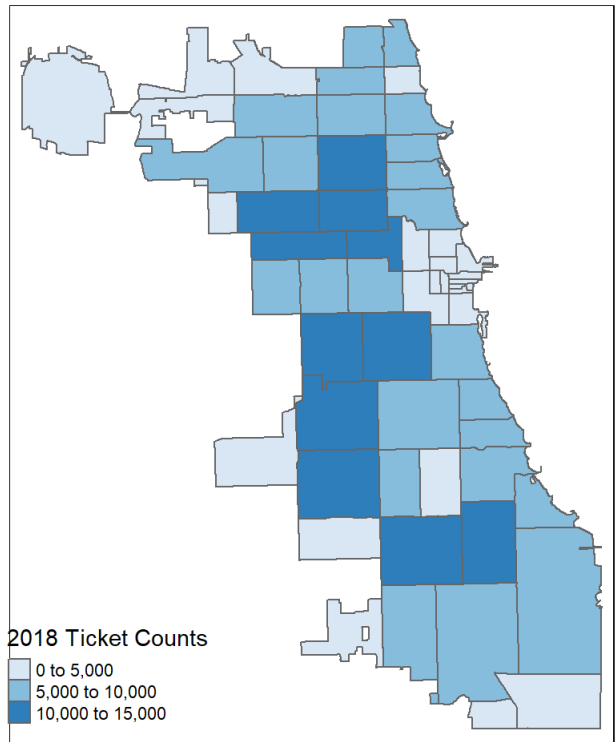

Figure 7. Zonal Distriution of Ticket Counts in 2018

The 'license_plate_state' variable explains which state the license plates belong to, figure 8 displays the ratios of Illinois license plates getting parking tickets in each year. The ratios are around 90%, suggesting most of the parking violation cases in Chicago are caused by local cars, thus the relationship concluded from linear regression method is representative, though there may exit a bias since the socio-economic data out of Chicago area is not considered.
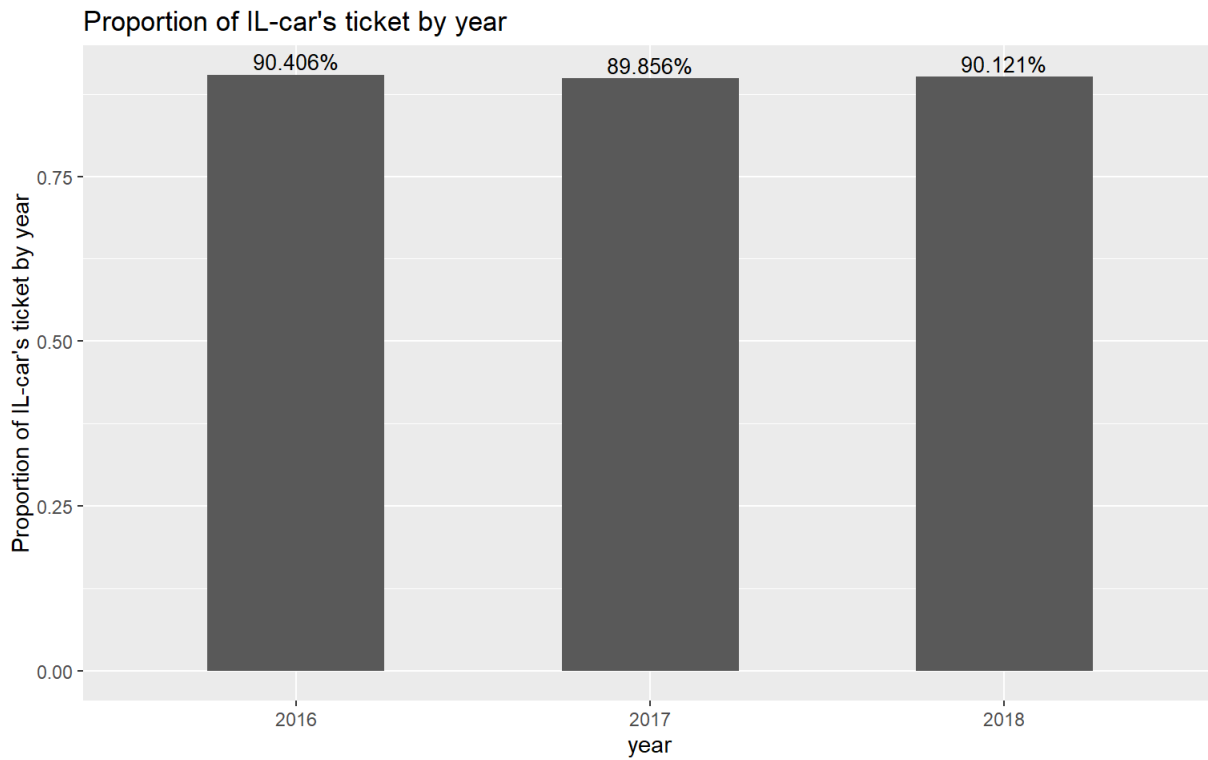
Figure 8. Ratio of Illinois License Plates in Parking Tickets

3.2.2 Linear Regression on ticket counts

Based on the AIC criteria, the independent variables of the linear model after selection are population, median age, mean family income, and 3 or above car ownership per household. The linear model could be used in visualizing the correlation after check the validity of the model.
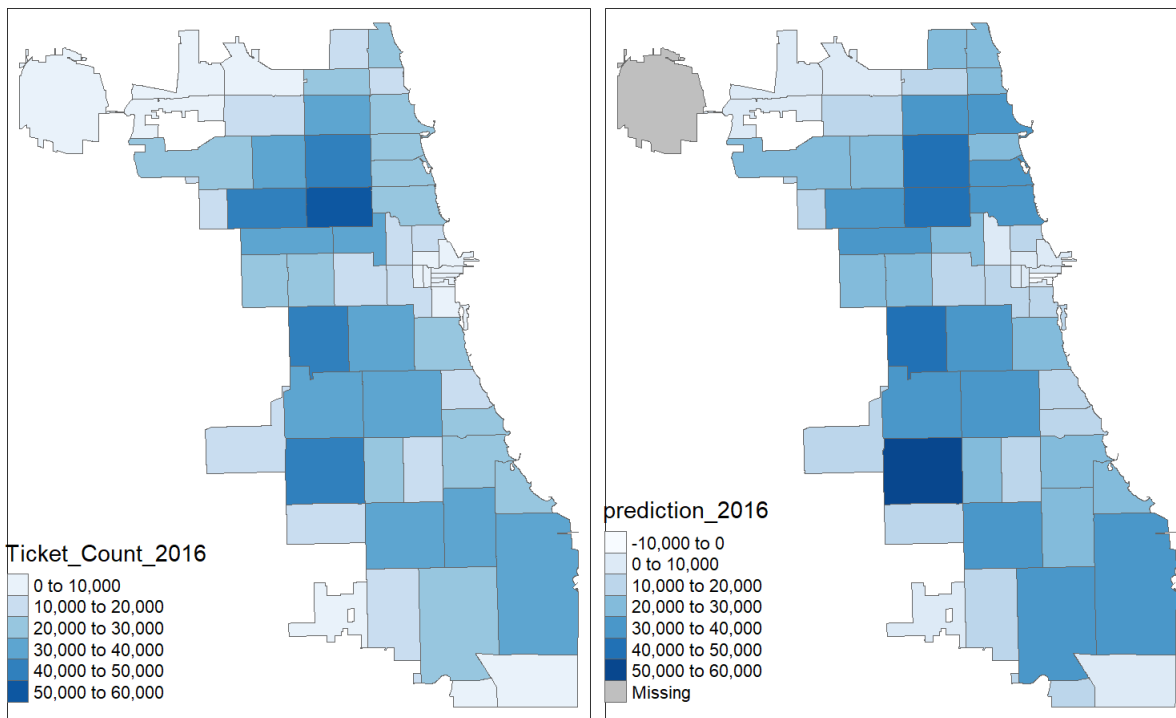

Figure 9. Actual and Predictive Zonal Distriution of Ticket Counts in 2016

The trained linear regression model is,

Ticketcount = 1.33e+04 + 5.45e-01 $Population$ - 3.39e+02MedianAge - 2.328e-02 $FamilyIncome - 3.556 + 00$ Vehicle3

3.2.3 Performance of the trained linear regression model

The residual plot of the new model shows that error doesn't have obvious bias, and is around 0. At the same time, the Q-Q plot shows that the normal hypothesis is true since the points all distribute around y=x.

Residuals vs Fitted

lm(Ticket_Count ~ population + Median_Age + Mean_Family_Income + Vehicle_3)

Figure 10. Residual Plot of Trained Linear Regression Model

Normal Q-Q

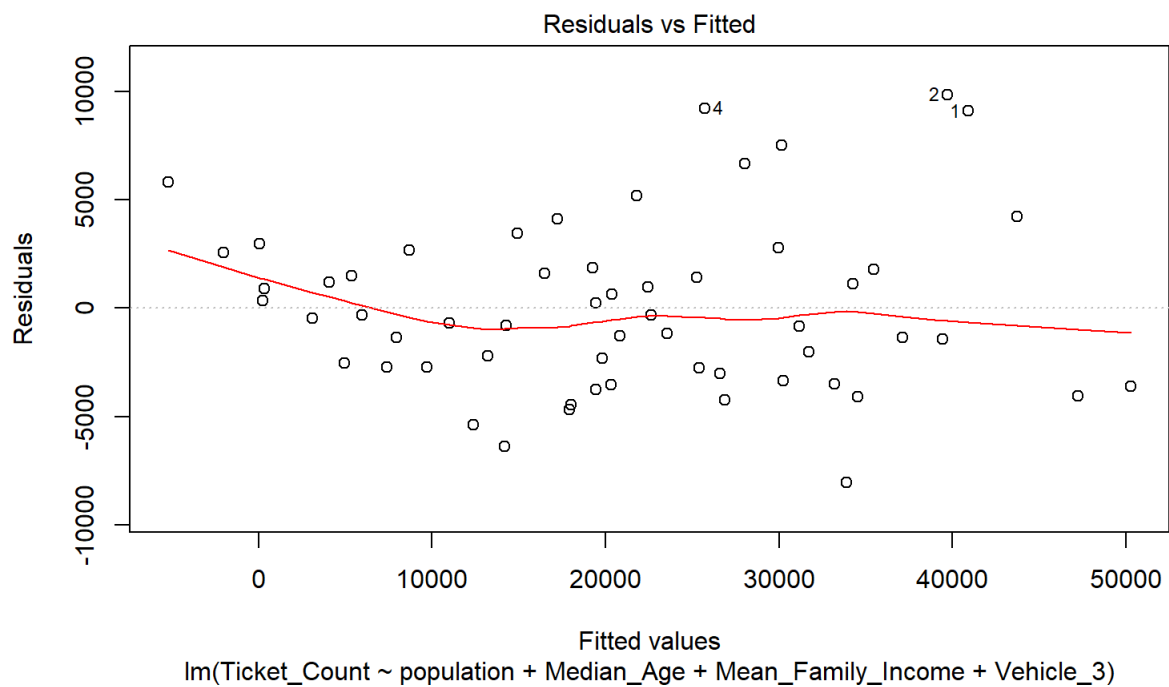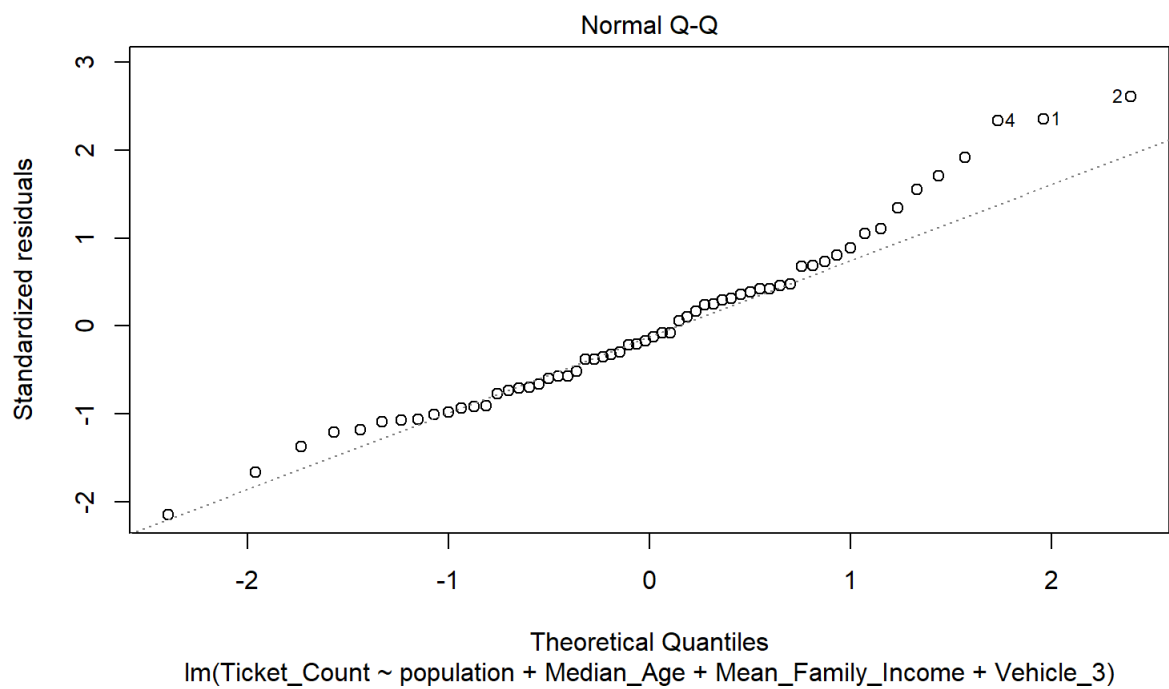lm(Ticket_Count ~ population + Median_Age + Mean_Family_Income + Vehicle_3)

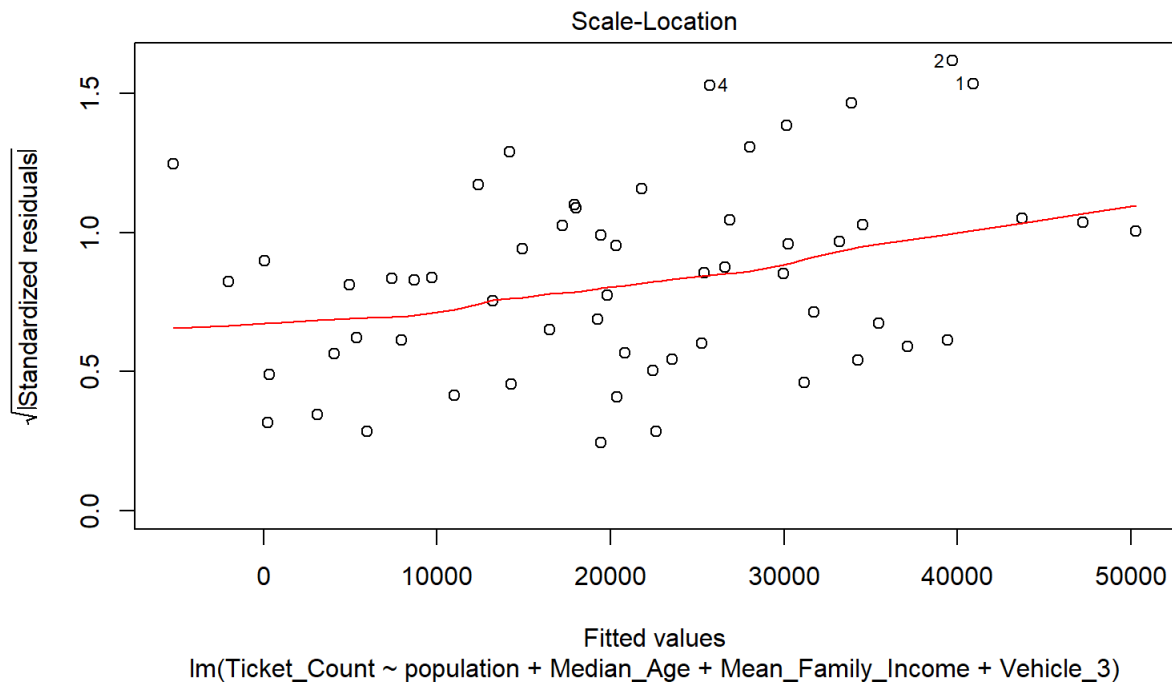Figure 10. Residual Plot of Trained Linear Regression Model

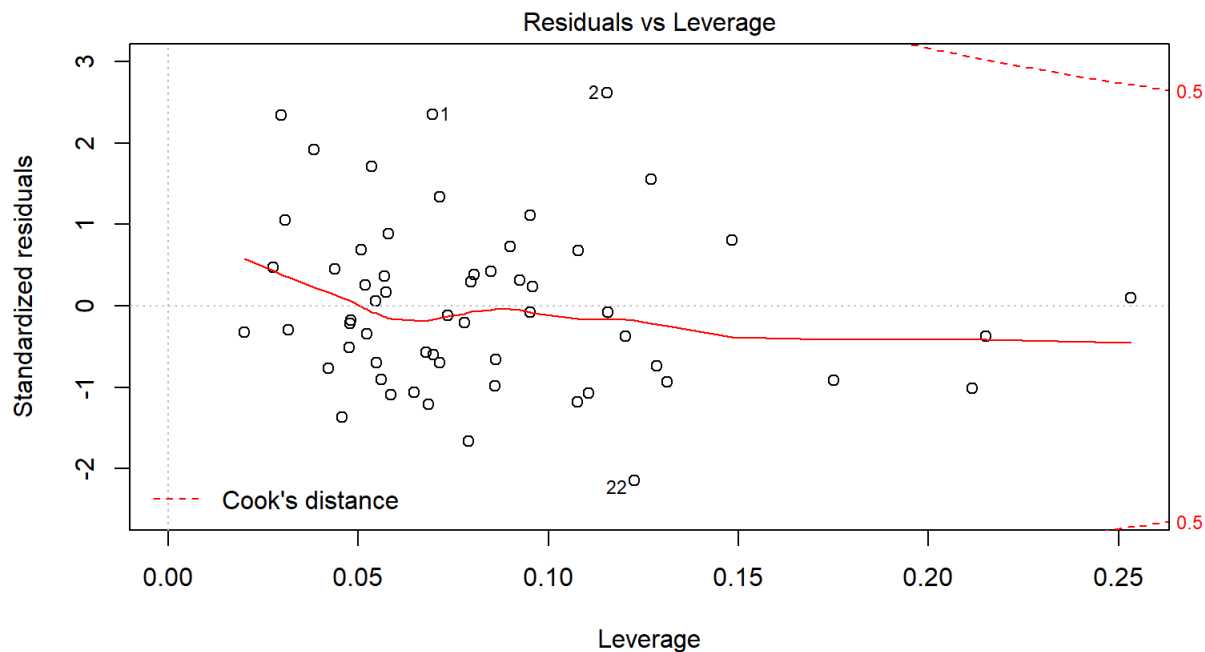Figure 10. Residual Plot of Trained Linear Regression Model

Figure 10. Residual Plot of Trained Linear Regression Model

Since the p-value or the F-value evaluates the variables in the model collectively. As it is shown in the anova table, the p-value of mean family income is 0.05 and so is the population. Median age, and 3 or above car ownership per household are all far less than 0.05, which means statistically significant to the model.

Table 1. Anova Table of Trained Linear Regression Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
population	1	9205699700	9205699700	574.520975	0.0000000
Median_Age	1	316100535	316100535	19.727603	0.0000436
Mean_Family_Income	1	63596762	63596762	3.969027	0.0513176
Vehicle_3	1	344445297	344445297	21.496579	0.0000223

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	55	881279371	16023261	NA	NA

For the selected linear model, the multiple R-squared is 0.9185, and the adjusted R-squared is 0.9126. Results show the model is in a good fitting condition.

3.3 Trend of Bankruptcy Ralated to Parking Tickets

The vehicles that had parking ticket in 2016 and 2017 were separated into 4 groups by the number of violations. Figure 13 shows that 70% of the vehicles had more than three tickets within these two years and only 20% of the vehicle had just one violation within these two years.
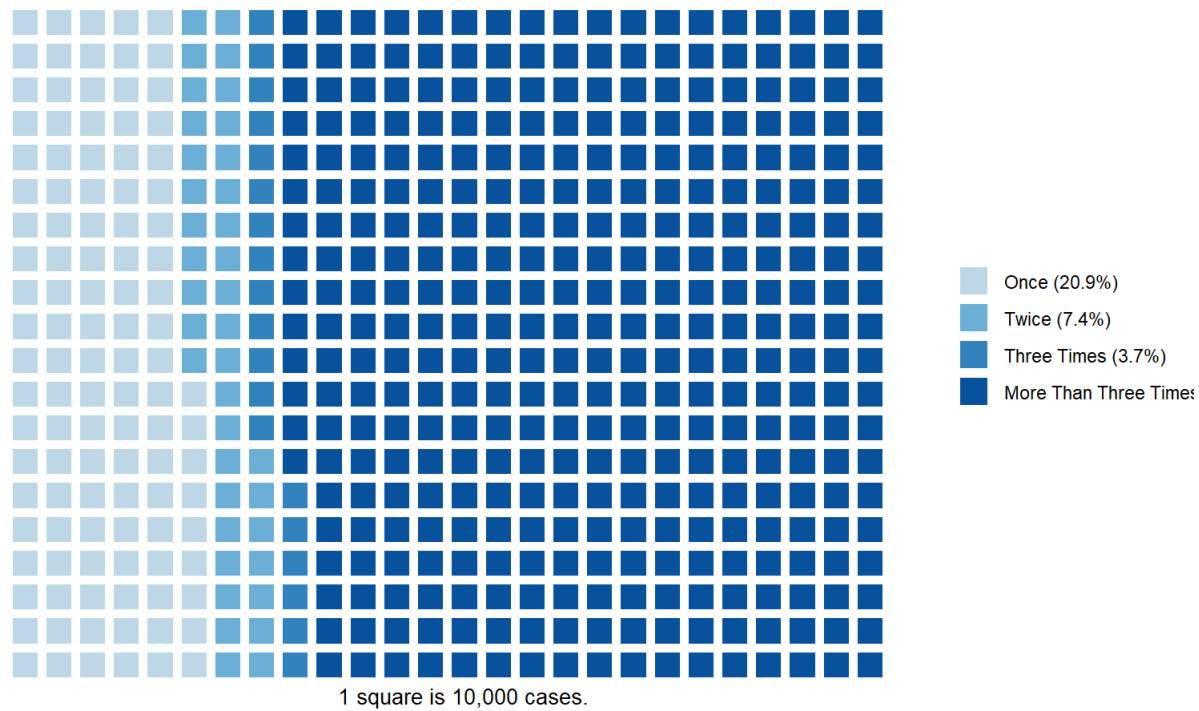

Figure 11. Waffle Chart of Different Violation Times

In this study, the vehicles with more than two times violation within these two years are the subject for bankruptcy analysis because if the vehicle only had one parking ticket in two years, it would be unlikely the owner of the vehicle to apply the Chapter 13 bankruptcies. By plotting violation reason for all these cases in different year, it is clearly to see that expired plates or sticker violation is the top reason for vehicle to get a ticket for all three years which is consistent with the statement in Curtis Black's article that the main source of the parking ticket debt is the tickets related to expired stickers (Black, 2018).

Figure 12. Top 5 Violation Types

By tracking the ticket queue status in the data, 497 people claimed bankruptcies after receiving several tickets in 2017. The number was reduced by 55% compared to 1114 people who claimed bankruptcies in 2016. Also, there were 1146 bankruptcies cases that changed to non-bankruptcies status in 2017 and it was a 16% increase compared to 986 cases from bankruptcies status to non-bankruptcies status in 2016. The trend from 2016 to 2017 is not consistent with the general trend that number of bankruptcy cases raised from 1000 to 10000 from 2007 to 2017 mentioned in Sanchez and Kambhampati's paper (Sanchez & Kambhampati, 2018).
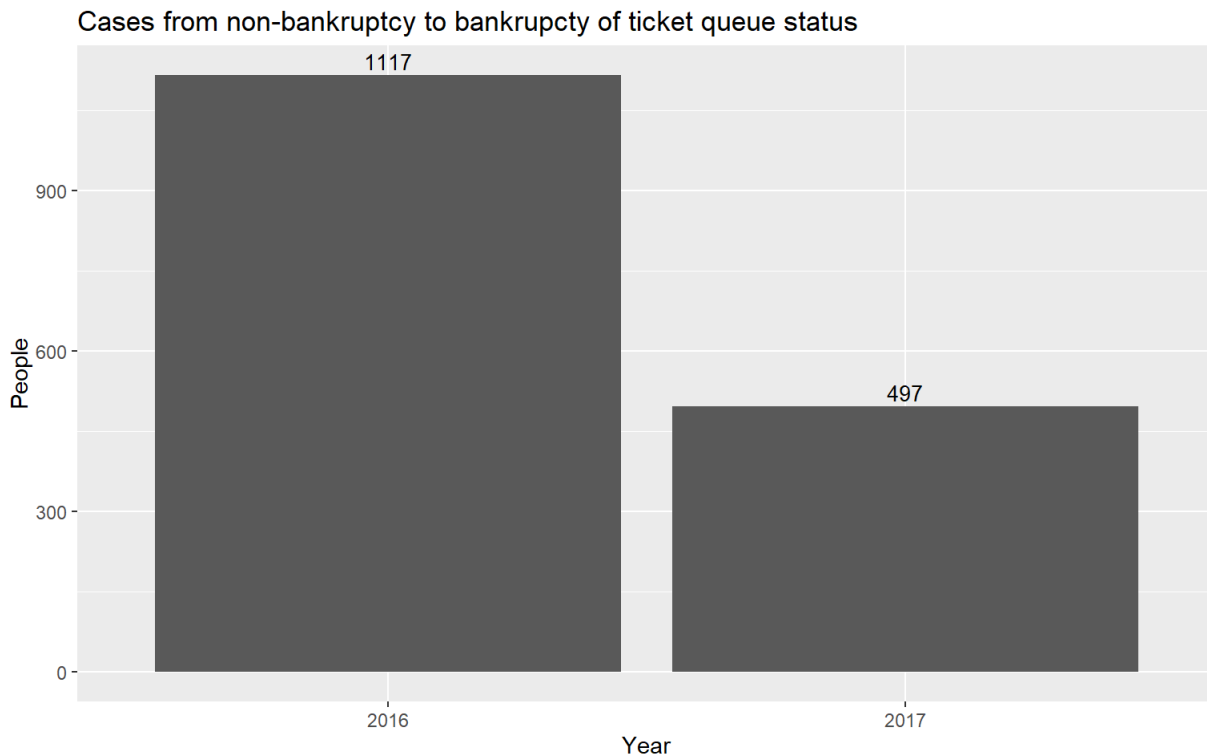

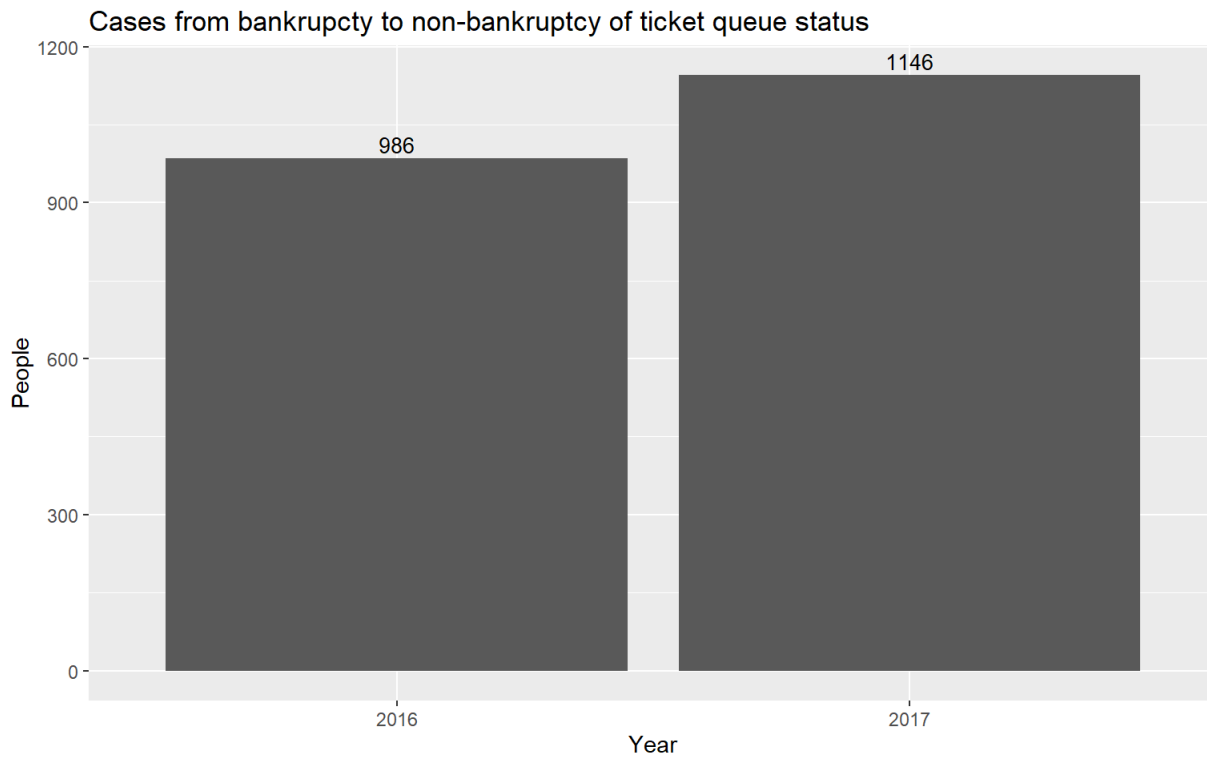Figure 13. Counts of Bankruptcy to Non-bankruptcy cases

Figure 14. Counts of Non-bankruptcy to Bankruptcy cases

One of the reasons is that since 2016, the government unveiled a new law that the government would hold the liens of impounded cars until the underlying ticket debt getting paid out fully during the bankruptcy payment plan (Sanchez & Kambhampati, 2018). Thus, filing Chapter 13 bankruptcy to release the impounded cars or reinstate the suspension licenses would not be a feasible method anymore, which means it is possible that the government could retrieve back more debt with this new law.

4 Conclusion and Discussion

Summary

1. Both number of tickets issued and total fine amount decreased by about 3% from 2016 to 2017. There was 28.6% drop on total fine amount from 2017 to 2018 by only apply first four months data of these two years.

2. February is the month that has the lowest total counts of ticket and total fine amount for 2016, 2017, and 2018.

3. Parking tickets are issued 75% less in weekend compared to in weekdays. The majority of the tickets issued was from residential area during the weekend and the majority of the tickets issued was from both residential area and CBD during weekdays.

4. Chicago area are divided into Transportation Analysis Zones by zip codes, linear regression is utilized to generate the relationship between parking tickets count with socio-economic variables, the 4 most significant ones are selected by AIC criteria.

5. More than 80% of the people got tickets more than once in a year. The top violation reason was expiration of plate or stickers.

6. From 2016 to 2017, the number of people claim bankruptcy decreased by 55% and the number of bankruptcy case that change to non-bankruptcy status increased by 16%

Critique on the methods

1. The linear regression model between the parking ticket and the social econometrics is not solid enough. First, since the parking tickets could be reverent with the attributes of social patterns, it could also be effects by the other significant factors including spatial factors. For examples, as for the Chicago downtown area where the parking demand is enormous, the supply side could be limited by early urban planning. The parking violation would be more likely to happen. By just using the social econometrics dataset, though it might be suitable for relatively rural area, the application of the same kind of regression on all zip codes area. Second, the results could have been more persuasive if using dependent variable per capita. Intuitively speaking, the total violation count will be increasing with population.

2. Bankruptcy As for the discussion on bankruptcy, the only method used is that to search for people that went into bankruptcy. Among these people, many could choose to go into bankruptcy to avoid tickets' debt. Only using these parts of data would not be enough if the purpose is to focus to the group of people that really cannot pay their ticket debt.

3. Continuity The result from the temporal analysis is not being used in the latter part. It could have been better if the results from temporal analysis applied to the spatial and bankruptcy analysis.

Future work

1. Did a regression based on violation per capita and the results were not showing strong correlation. Investigate the reasons for that is needed to be done in detail.
2. Social econometrics from other year beside than 2016 is needed in order to limit the randomness of one year data.
3. A method to select the real bankruptcy people and analysis on them is needed to help optimizing policy.

5 References

Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.

Brockway, Mike. "$1.5 Billion in Unpaid Tickets Could Be Huge Cash Cow for Chicago." DNAinfo Chicago, 31 Mar. 2015, www.dnainfo.com/chicago/20150331/downtown/15-billion-unpaid-tickets-could-be-huge-cash-cow-for-chicago.

"Chapter 13 Bankruptcy and Unpaid Chicago Parking Tickets." DebtStoppers, Bankruptcy Law Firm Robert J. Semrad & Assoc., LLC, www.debtstoppers.com/eliminating-debt/chapter-13-bankruptcy-and-unpaid-chicago-parking-tickets/.

Sanchez, Melissa, and Sandhya Kambhampati. "How Chicago Ticket Debt Sends Black Motorists Into Bankruptcy." ProPublica, 27 Feb. 2018, features.propublica.org/driven-into-debt/chicago-ticket-debt-bankruptcy/

Thomas, Steven. "How Much 25 Major Cities Make in Parking Ticket Revenue Per Capita." TopView Sightseeing, 2018, www.topviewnyc.com/packages/how-much-25-major-cities-make-in-parking-ticket-revenue-per-capita.