**Crime analysis in Washington DC**

Jiachi Zhang, Xiaosong Wang, Qi Xie, Zheyi Wang

http://crime2017.georgetown.domains/uncategorized/crime-analysis-in-washington-dc/

**Introduction**

There is always significant concerning on the reasons that can influence the number of juvenile crime, since juvenile crime can touch millions of people’s life every year. In 1997, law enforcement officials arrested 2.8 million people under the age of 18, accounting for one in five of all arrests that year. [1] And many research have focused on how the environment of their living place. For example, The LA’s BEST after-school program demonstrates statistically and substantively positive effects on youth crime abatement, especially for students who attend at least 10 days per month. [2] So we assume that the crime situation can provide massive influence on juvenile crimes. Besides, we use a variable “Safety” which relates to number of crime and how severe they were to describe the crime situation of each location. With the “Safety”, we can analysis the relationship between juvenile crime and crime situation.

**Background**

As we known, crimes can cost our society dearly and threaten our safety in our daily life [3]. And the influenced juveniles can make the situation worse. For the society, the distribution of crimes can help police departments deploy their men to prevent the crimes or add monitors at the best location so that they can use their funds more efficiently. Besides, according to some research, after school projects can reduce juvenile crime, [2] we can assume that because juveniles can easily be influence by the environment they live, their crime rate can relate to the crime rate of the place they live.

Historically, the work of dealing with crimes is the prerogative for the criminal justice and law enforcement specialists. However, the development of computer science and data science can help the law enforcement officers and detectives solve the problems by using algorithms in data analysis [3].

**Data set and Analysis**

Data description:

We gathered three datasets from <http://opendata.dc.gov/datasets>. They are crime data in 2017, juvenile data from 2011 to 2017, felony data from 2011 to 2016. In these data:

“SHIFT” is when the crime happens (day, night or midnight). We can see that crimes are more likely happen in the evening than in the daytime

“METHOD” is how the crime is committed (with a gun/knife or other way). From the above results we can conclude that most of reported crimes do not involve with weapons.

“OFFENSE” refers to crime offenses which includes many types of crime. But from the results most of crimes belong to a broad inclusion of Theft offenses including embezzlement, theft of services and fraud/false pretenses.

The first number in “PSA” can be used to get the location. And the data divided D.C. into 7 areas. Like the picture shows:

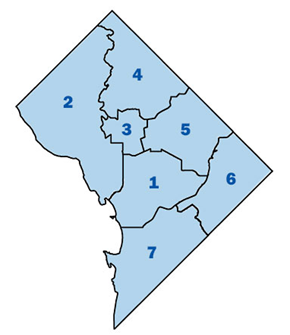


Figure 1 areas of D.C.

Data Pre-processing:

a) drop the attributes which is not helpful for our further questions

Since the object\_id is a unique code for each observation and it is not helpful for us to do the association rule mining or machine learning. We decided to drop the attribute.

b) encode some nominal attributes or convert them to numeric attributes

For attribute “OFFENSE”, we encode this using number 1-9, where the bigger number indicates the higher level of crime.

And for some other attributes such as “SHIFT” or “METHOD”, we map them to do the following machine learnings.

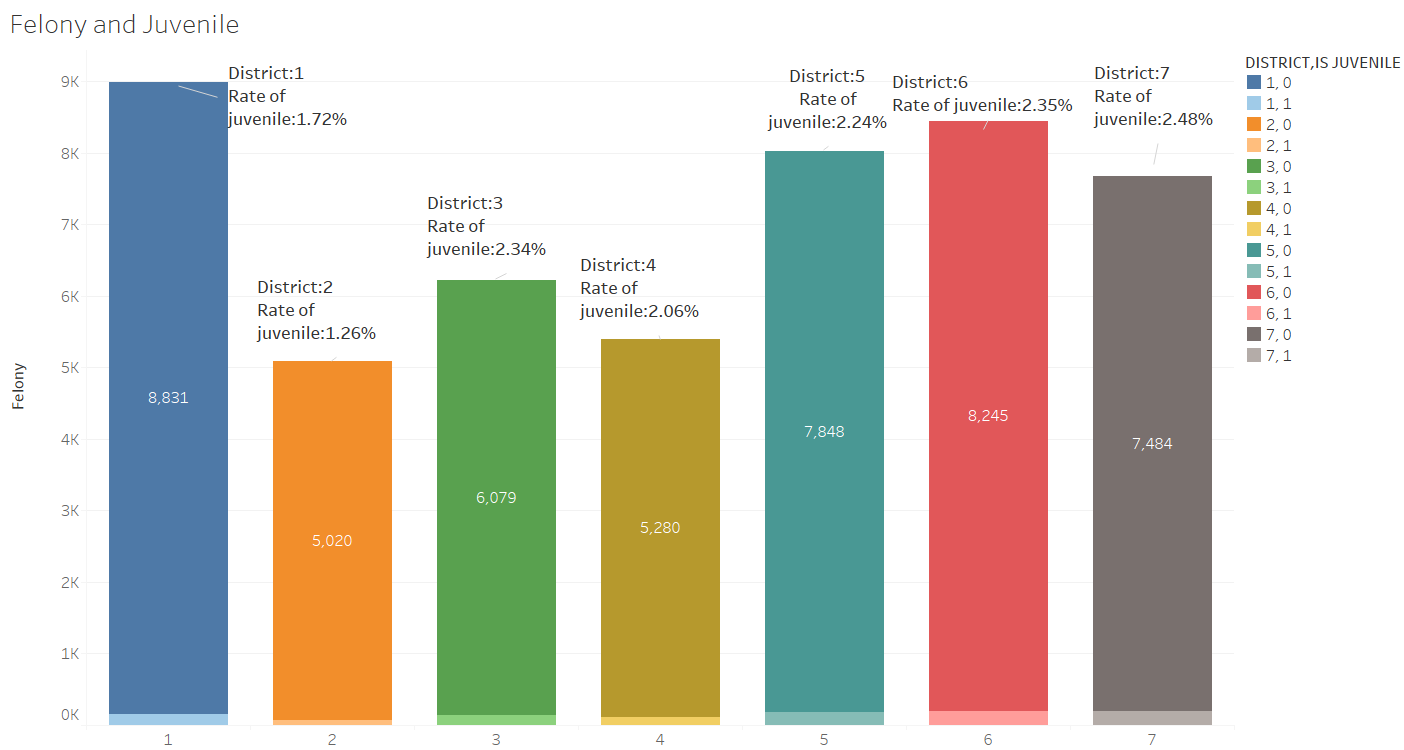
New variable “Safety”:

To describe the crime situation in an area, we created a new variable “Safety”. “Safety” is related to “OFFENSE” and the number of each kind of “OFFENSE”:

Variable *oi* means the weight of each offense ('THEFT/OTHER': 0.1, 'THEFT F/AUTO': 0.1, 'MOTOR VEHICLE THEFT': 0.2, 'BURGLARY': 0.5, 'ARSON': 1.0, 'ASSAULT W/DANGEROUS WEAPON': 0.9, 'ROBBERY': 0.7, 'SEX ABUSE': 0.7, 'HOMICIDE': 0.5). The more serious the crimes are, the bigger their weights are. And *ci* means the number of each offense. k is a constant to adjust the Safety in a reasonable range.

**Visualization & Analysis(part)**

(Full version: http://crime2017.georgetown.domains/uncategorized/analysis/)

****

Here’s a graph about the rates of juvenile felony of each district from the felony data . Corresponding to the safety value of the districts, we can also find the relationships between juvenile felony and safety.

In the felony data set, we focus on the suspects in the data and we regard those suspects who were less than 20 years old as juveniles. Generally the rates are not big values, all of them are in the range of 1% to 3%, but the difference among them can also reflect the impact of safety on juvenile crime.

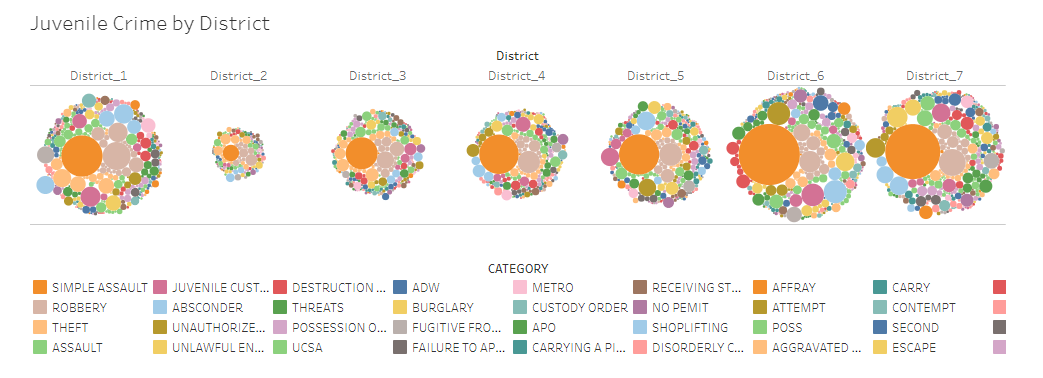
We notice that in district 2,4 and 1, the rates of juvenile felony are lower than the other districts. According to previous analysis, these 3 districts are also considered “safer” than other districts, especially district 2 which is the “safest”, has the lowest rate of juvenile felony which is only 1.26%. The other 4 districts all have a rate more than 2.2%.

In the other words, from this graph, relatively, district with a lower safety value which means it is safer also has a lower rate of juvenile felony. We may infer that a district with less felonies will play a positive role on prevent juveniles from crime.

****

As what we can see from the word cloud above, the word “Simple Assault” is prevalent in the data set, followed by the word “Robbery”, “Absconder” and “Juvenile Custody Order”.

We tried to find the topic “Theft” because it is dominant in general crime. But in juvenile crime cases “Theft” is not that significant. Why “Simple Assault” is prevalent instead of “Theft”? We think the main reason is that young people are tending to be aggressive when they’re getting angry.

****

This graph tell the details about the type of juvenile crime in different districts. We can see that those topics seem to have similar distribution in all these different districts.

According to the analysis, we can draw the conclusion that there is no obvious similarities between type of juvenile crime and general crime.

**Conclusion**

According to the statistics of these three datasets, we can get:

|  |  |  |  |
| --- | --- | --- | --- |
| District | Safety | Number of Juvenile crime | Rate of Juvenile felony (%) |
| 1 | 6.8096 | 1133 | 1.72 |
| 2 | 6.6511 | 219 | 1.26 |
| 3 | 6.8348 | 462 | 2.34 |
| 4 | 6.6849 | 641 | 2.06 |
| 5 | 6.9890 | 781 | 2.24 |
| 6 | 7.0959 | 1106 | 2.35 |
| 7 | 6.8433 | 1156 | 2.48 |

The safety shows us that district 2 is the safest area in D.C. and district 6 is the most dangerous area. What conform to our prediction is that the number of juvenile crime and rate of juvenile felony is the smallest. However, numbers of juvenile crime in district 1 and 7 are more than we expected. Besides, the rate of juvenile felony is higher than we expected too. Maybe we get this result because we didn’t collect the population and the density of population in each area. But from the result we get, we can make a conclusion that if an area is safe or have less crimes, the number of juvenile crime will be small. Besides, according to Ludwig, J., Duncan, G., & Hirschfield, P. (2001), which shows the offer to relocate families from high- to low-poverty neighborhoods reduces juvenile arrests for violent offenses by 30 to 50 percent of the arrest rate for controls, [4] strength the result we got.

Although some number is not fit with our prediction perfectly, we still can say that juvenile crime situation gets worse when the area is not safe approximately.

**Reference**

[1] Jacob, B. A., & Lefgren, L. (2003). Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime. *American Economic Review*, *93*(5), 1560-1577.

[2] Goldschmidt, P., Huang, D., & Chinen, M. (2007). The long-term effects of after-school programming on educational adjustment and juvenile crime: A study of the LA’s BEST after-school program. *Los Angeles: UCLA/CRESST. Retrieved September*, *8*, 2008.

[3] Nath, S. V. (2006, December). Crime pattern detection using data mining. In Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 ieee/wic/acm international conference on (pp. 41-44). IEEE

[4] Ludwig, J., Duncan, G., & Hirschfield, P. (2001). Urban poverty and juvenile crime evidence from a randomized housing-mobility experiment. The Quarterly Journal of Economics, 116(2), 655–679. https://doi.org/10.1162/00335530151144122