

Dallas Crime Incident Data Analysis

Aniruddha Perumalla – Data Science Intern – Techwave

Advised by Satya Puppala and Vinisha Gambali

July - August 2016

1 Background

The data that we have analyzed has officially been made publicly available by the Dallas Police Department (DPD) on [this website](#). It details various demographic- and bookkeeping-related information on an incident-by-incident basis as noted by the DPD. (A small sample of the types of dimensions noted: the demographics of the victim of the crime, the specific street address where the crime occurred, the type of crime, the specific date and time of the crime, the ID of the police watch area where the crime occurred, the city division where each crime occurred, etc.)

2 Methods

First, we performed exploratory analysis of the data through visualization in Tableau. Next, we ran the Apriori algorithm on the data, and examined the resulting rules to observe patterns in the data. Finally, we inspected the data from a geospatial stance, looking for insights with relation to zip code.

3 Insights

We discovered the following insights through exploratory analysis in Tableau. The graphs from which we drew these conclusions can be found at [this Dropbox link](#) (open the link, then click “no thanks, continue to document”).

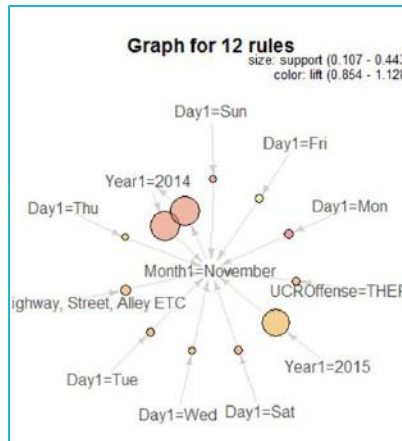
- The number of incidents *always* rises (spikes) on the first day of each month, regardless of the the number of incidents of the last day of the previous month.
- Ross Bennett (5741), Forest Audelia (5449), Central CFHawn (5166), CampWisdom Chaucer (5140) are the most active TAAGs*.
- Very few drug-related crimes (3) take place; a small amount of hate-related crimes (33) take place; zero hate- *and* drug-related crimes take place.
- [Black females] (23096), [Latino males] (22536), and [White males] (22533) are the most common victims.
- There are far more [Black female] (23096) victims than [Latino female] (17256) or [White female] (17186) victims.
- There are more [Latino male] (22536) / [White male] (22533) victims than [Black male] (19373) victims.
- The most active victims not from Dallas are from: Garland (2196), Mesquite (1958), and Irving (1877).
- The most active community by far is Northwest (4664), followed by Forest Audelia (2727); the least active community is Pemberton Hill (181).
 - Specifically, in both the Northwest Community and the Forest Audelia Community, the most active premises are Apartment Parking lots (905 / 759), Apartment Residences (501 / 674), and Apartment Complexes/Buildings (452 / 357).
- 1400 S Lamar Street (767) and 8687 N Central Expressway (607) are the most active addresses.

* Targeted Area Action Grids (TAAGs) are areas with high crime rates independently outlined by the Dallas City Government. They have no relation to any county, block, or street names.

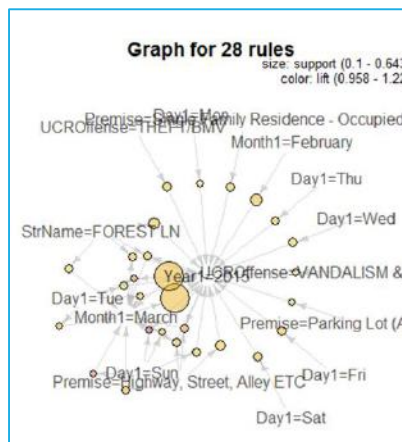
- The number of crimes from Jan 2016 to Jun 2016 (6 months) is roughly equal (53k) to the number of crimes from May 2014 to December 2016 (8 months).
- The number of crimes in 2015 (93k) is roughly 75% larger than the number of crimes in the period from May 2014 to Dec 2014, even though there are only 33% more days in the whole of 2015 than in the period from May 2014 to Dec 2014.
- Northeast (34729) is the most active division; North central (20917) is the least active division.
- Nearly 85% of all incidents are investigated further as a follow-up.
- Of the victims who were involved with drug usage, [Black males] (303) are the most numerous, followed by [White males] (212) and [Latino males] (203).
- The most common description for crimes is Theft (50733), followed by Burglary (28414) and Criminal Mischief / Vandalism (25589).
- 60% of the crimes are due to Burglary, Theft, Crime Mischief and Robbery.
- The most common street for crimes is Forest Lane (167), followed by Harry Hines Boulevard (131).
- The vast majority of cases are suspended (i.e. perpetrators are paroled) (171488 out of 198841 = 86%).
- The peak timings of crimes are around 5:00 PM. Fridays and Saturdays are the peak days.
- Most accident prone areas are: between 5500 LB J Freeway and 10000 LB J Freeway (specifically 9700 LB J Freeway), and 1700N Central Expressway.

We then tried to find patterns in the data by using the Apriori algorithm, which finds “rules” that describe links between certain dimensions. We ran the Apriori algorithm on the five zip codes where the most crimes occurred. The zip codes we chose were the following: 75217, 75243, 75228, 75216, and 75220. It should be noted that all of these zip codes are not truly in the inner city, but rather a bit farther outside the inner city zone. We chose to operate on the following most significant dimensions: UCR Offense, Zip Code, Street Name, Incident Number, Premise, Day, Month, and Year.

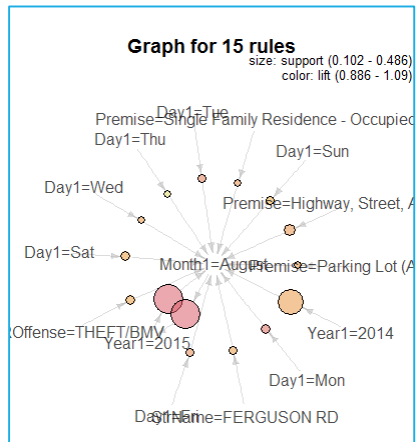
We chose to group the Apriori results by zip code, since this would allow us to more easily present the findings. The corresponding rules visualization produced in R accompanies each finding (by zip code) below. We discovered the following insights for each zip code:



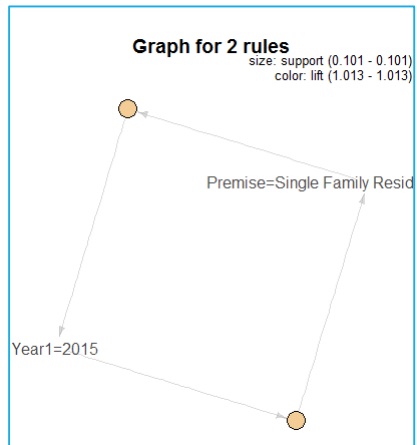
- ZC 75217
 - It can be observed that for this zip code, many thefts occurred on transportation pathways in both November 2015 and November 2014.



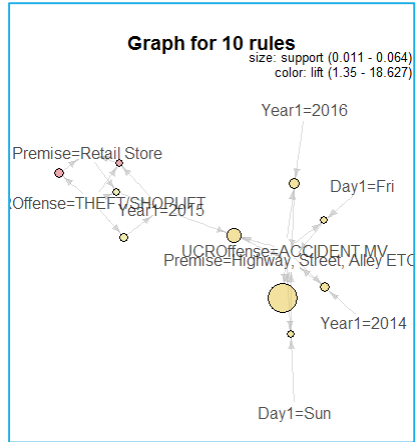
- ZC 75243
 - The rules in this zip code mainly center on crimes that occurred in March 2015. First, in March 2015, there were many crimes on transportation pathways. One of these pathways, Forest Lane, seems to have been the location of several occurrences of vandalism and other criminal mischief. Moreover, we can conclude that most of the crimes occurred on weekends.



- ZC 75228
 - August 2014 and 2015 seem to have been significant times for crimes in this zip code. Many crimes occurred in single-family households and transportation pathways such as Ferguson Road. Furthermore, we can state that a large number of BMV (Bureau of Motor Vehicles) –related crimes occurred in apartment parking lots.



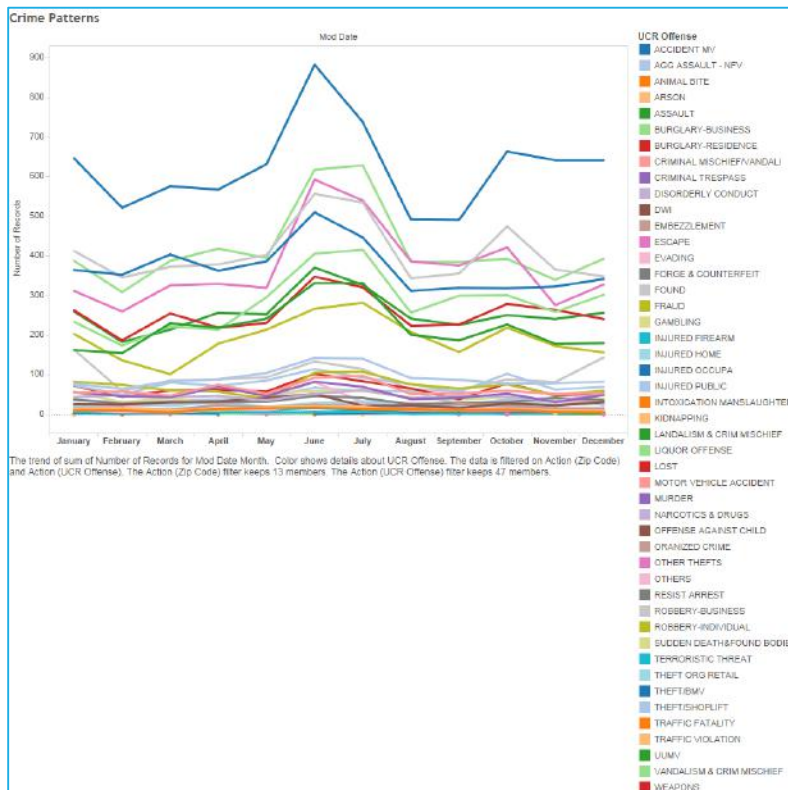
- ZC 75216
 - This graph is the simplest to read; we can easily observe that for this zip code, there were a significant amount of crimes in single-family households in the year 2015.

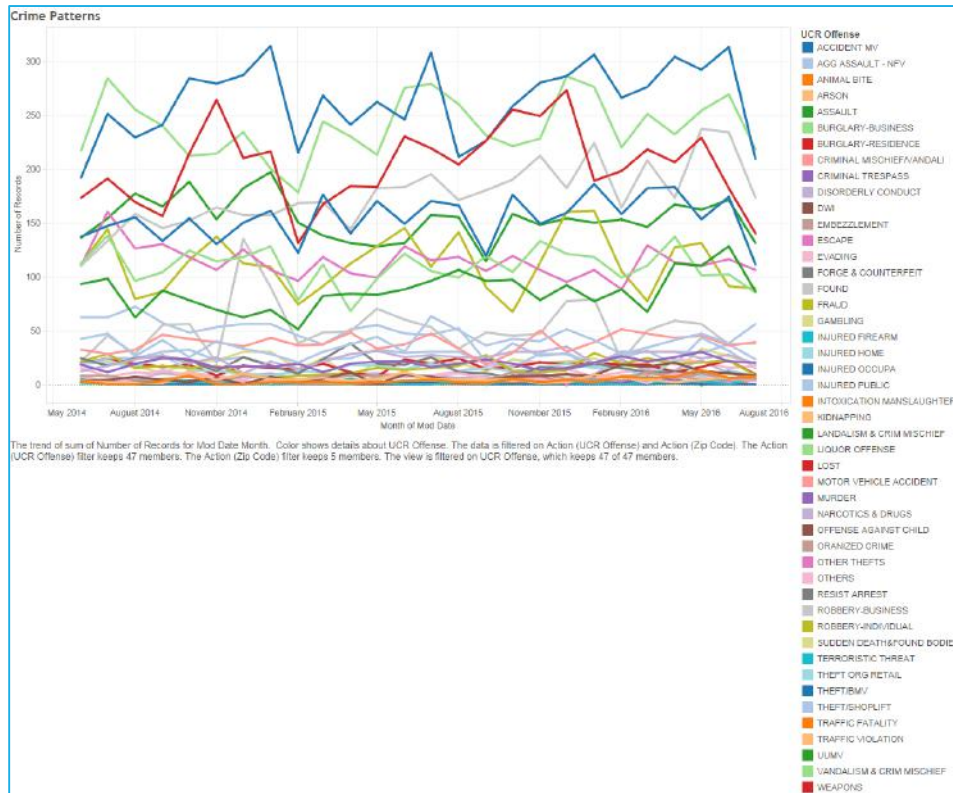


- ZC 75220
 - Many motor vehicle accidents occurred in this zip code in 2014 and 2016 on transportation pathways. Additionally, several thefts and shoplifts occurred at retail outlets in 2015.

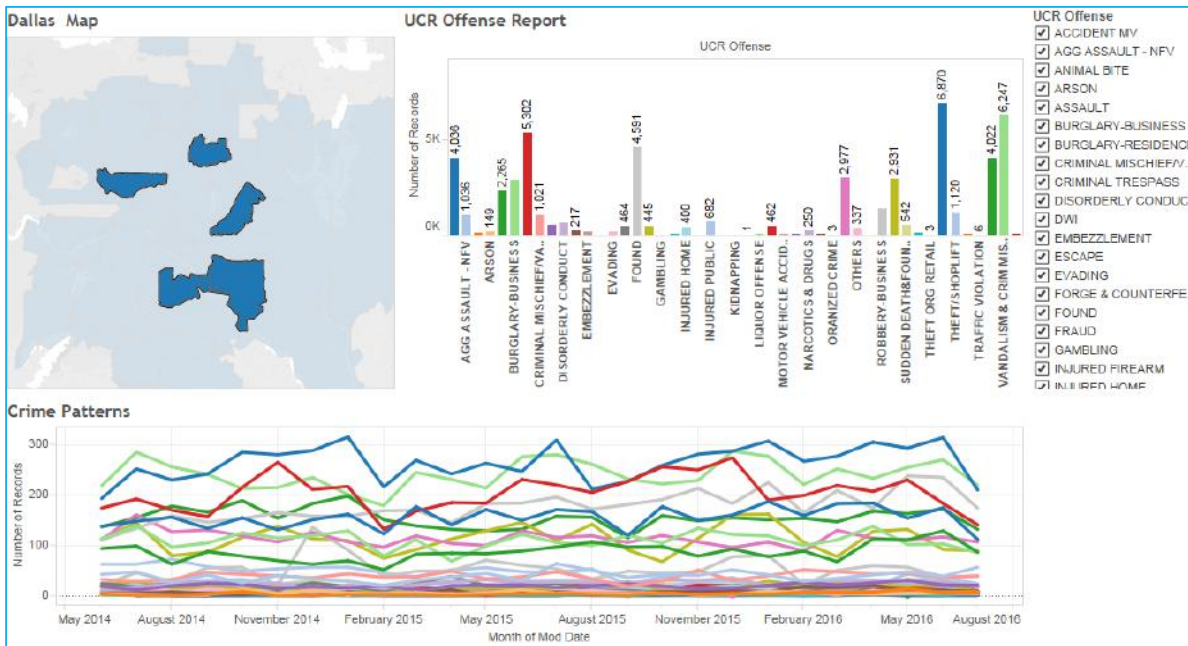
Finally, we examined the data from a geospatial perspective; that is, we performed more exploratory analysis through Tableau visuals, in which zip code was the distinguishing variable.

Graph A:
 Month of
 Crime
 vs.
 Number
 of Records
 (by
 Offense)





Graph B:
Specific
Month of
Crime in
the top 5
Zip Codes
vs.
Number of
Records
(by
Offense)



Graph B: Visuals for the Top 5 Zip Codes

Through this geospatial viewpoint, we determined the following from the graphs shown above.

- The zip codes where the most weapon-related crimes occur are: 75215, 75216, 75217, 75227, 75228, 75237, 75241, 75220, and 75243. Most of these zip codes are surrounding the city center rather than directly in the city center.
- Most crimes occur in zip codes close to the center of the city, rather than the outskirts.
- Assault crimes tend to peak in the autumn (August-November).
- Robberies tend to peak in the early summer (May).
- Found crimes tend to peak in the autumn. This phenomenon can especially be visualized for the zip code 75210.
- Thefts in zip codes close to the center of the city tend to peak in the autumn.
- In general, crimes in zip codes close to the center of the city peak in June.

4 Prescriptives

We found that the Dallas Police Department would do well to increase surveillance of the Northwest community, since by far the most crimes occur there. In addition, the fact that a significant portion of crimes involve stolen vehicles needs to be addressed, perhaps through public awareness to lock cars before leaving them. Since many crimes occur near apartment complexes (specifically in parking lots), more lights could be installed near these apartments to increase visibility (and consequently, the chance of catching the perpetrator in the act), and “emergency kiosks” could be installed to immediately notify police rather than calling 911. “911” is the emergency number in the USA; dialing it will direct the caller to an operator, who will notify the police. Installing emergency kiosks would bypass the redirection step involving the operator, therefore making the process faster. Moreover, one of the more interesting insights we made was that of the “fresh opportunity spike” of crime activity on the first day of each month; the DPD would do well to take note of this and tighten monitoring on the first day of each month. It was interesting to note the gender and race differences among the victims, but it is not clear what needs to be done with regard to black females being the most common victims, or the gap between black female victims and female victims of other races.

From the Apriori results, we can see specifically that a significant portion of the crimes in the top 5 zip codes are occurring in transportation pathways. This may serve as a warning that police effectiveness regarding motor vehicle crimes is lacking. The thefts occurring in retail outlets are to be expected, but the DPD may want to look into those crimes in that zip code to investigate further. In addition, single-family households, who can be expected to live in standalone homes as opposed to apartment complexes, are under threat. The DPD may want to raise awareness about home security systems like cameras, triggered alarms, and home appliance automation.

With regard to the geospatial analysis, we see that (as expected) most crimes occur in the inner city areas and the areas just outside the city center. Moreover, these crimes that occur in these areas happen most often in the summer (June); it could be the case that criminals are more active in the summer than the winter simply due to more favorable weather conditions. (In fact, during analysis of another set of Dallas crime data, we found that the most crimes occur in warm weather.) The DPD may have to increase surveillance and attentiveness to crimes in the city center during the summer as compared to the winter.

5 Conclusion

Given incident-by-incident crime data for Dallas, we performed exploratory data analysis in Tableau, Apriori algorithm rule generation, and geospatial analysis. We found many interesting insights from visualization in Tableau, such as the crime victim distribution among whites, blacks, and Latinos and males and females. We also discovered relationships between crime rates and seasons and premises. The graphs returned by the Apriori analysis highlighted the high crime rate on transportation pathways and in specific months of the year. Finally, our analysis from a geospatial perspective produced several interesting observations regarding the geographic distribution of crimes in the inner city and suburbs.