Crime Occurrence Analysis in Chicago City

Author: Sai Krishna Vithal Lolla, Co-Author: Jiawen Liu
Management Information System, Oklahoma State University, Stillwater, OK 74078

Introduction

Metropolitan cities are often at the top of crime charts. Therefore, police departments in big cities want to understand, predict and if possible prevent or to mitigate potential damage from crimes. This project is aimed at analyzing crime in Chicago City.

The data set has been obtained from City of Chicago's website. It has crime records across Chicago over the past decade. Final data set used is a subset with even distribution across the time period. Data about offenses includes fields depicting the nature of crime, time of crime, block-level location information, legal treatment of crime and resulting punishment. This data set is then transformed for further analysis.

JMP® Pro 10's sophisticated graph building techniques are used in this project. The objective is to analyze the data and find possible patterns between crime rate and crime location. In addition to that, offense rate variation based on time is also explored. With such understanding, authorities can proactively take measures to prevent some of the potential crimes.

Methods

Step I Data Preparation

The data preparation primarily required consolidation by offense type. Similar offenses were attributed to multiple categories due to varying degrees of offense. These many categories impact precision of the analysis. This project varies from typical analytics projects in that most of the analysis results are graphical in nature. Using data with so many categories would dilute the results and mask most trends. Hence, all non-varying values in crime type variable were binned into broader categories.

Weekdays were extracted for each date value and populated into a separate field. These values of the resulting attribute would be used to check and compare crime occurrence across the days of a week.

There were fields such as offense identification numbers, landmarks for offense location, FBI codes and others which could lead to case identification and were removed. Furthermore, these fields were not really helpful for the current analysis.

Trimming and consolidation operations mentioned above resulted in transformed and standardized data set which was more analyzable.

[1] Categorical variable primary type is binned to reduce number of categories. These bins are number coded.

[2] Date variable is extracted into weekdays, months and years.

Step II Analysis using Graph Builder:

Graph Builder feature of JMP® Pro 10 software is used for analysis of the transformed data set. Out of the many graph types available in JMP® Pro 10, Histograms and Contour Graphs were used. These two types of graphs were chosen because of the nature of data and areas of analysis. For analysis of variables with finite numbers, Histograms are used and for analysis of high frequency fields, contour plots are used.

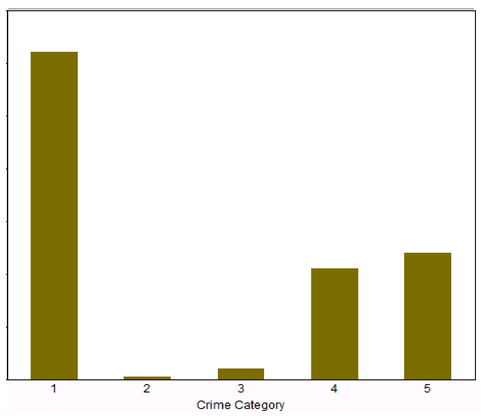

Figure 1.1

The crime category variable is set on X-axis. This gives us a split of crimes occurring across the various crime categories. For various analyses, other variables in the data set are brought onto the other axis, overlay, color and shape regions to generate relevant results.

In order to perform time-wise analysis of offenses, weekday variable and crime occurrence are considered. Similarly, to perform location-wise analysis of offenses, the latitude and longitude values of all crime locations are considered.

Graph Analysis

1. The split up of offense data among the various categories can be seen using a histogram. (Fig 1.1)
2. Day-wise occurrence analysis of offenses can be obtained by overlaying weekday on crime category. This results in multiple histograms, one graph each showing offenses per day of week. (Fig 1.2)
3. All the offenses recorded are divided by their latitude of occurrence. The total set of latitudes is divided into intervals and has the total range of longitudes repeating for each interval. Contour plots are used to efficiently display crime occurrence for each area across each crime category. The resultant graph shows the offenses recorded from each category, in the form of five intervals of latitudes. (Fig 1.3)
4. Time-wise occurrence analysis of offenses can be obtained using a contour plot . Time variable is brought in to Y-axis. The resultant plot captures the offense occurrence rate across various time periods. (Fig 1.4)

Results

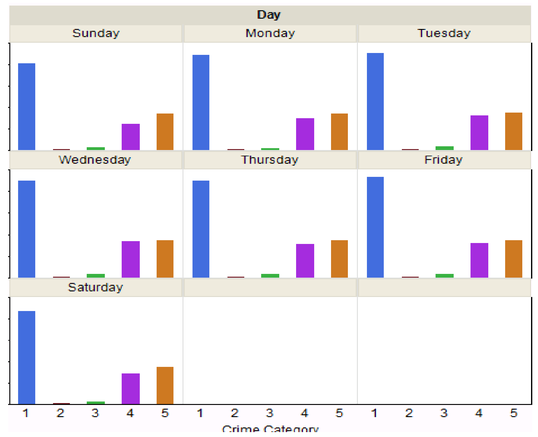

Figure 1.2

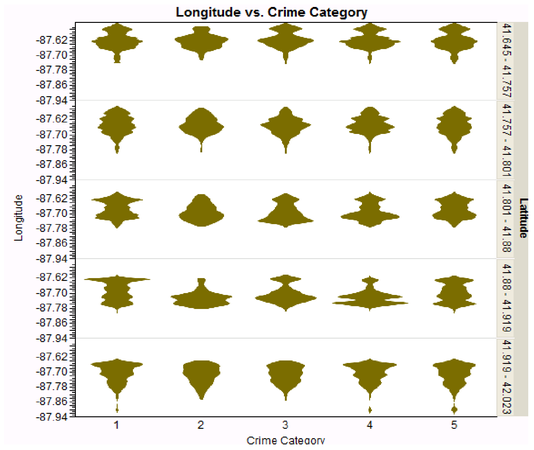

Figure 1.3

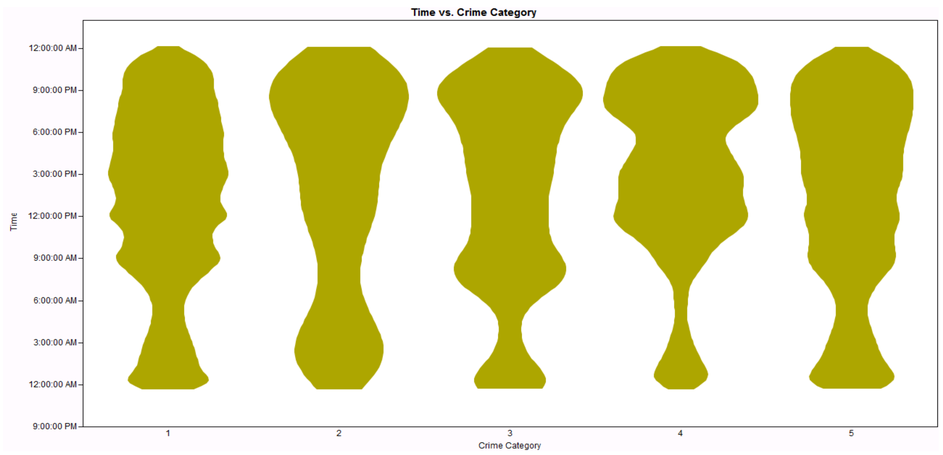

Figure 1.4

Discussion

Fig 1.1 shows crime category 1 (burglary, theft and robbery) seems to be the most frequent offense committed. It happens most on Friday and Saturday.

Fig 1.2 shows that most offenses seem to occur on Tuesdays and Fridays.

Fig 1.3 shows that type of crime offense varies across different areas. Burglaries (crime category 1) occur mostly in region 1. Homicide activity (crime category 2) occurs most in region 4. Sexual offenses (crime category 3) occur most in region 3. Drug abuse offenses (crime category 4) occur most in region 5. Other Criminal Damage (crime category 5) occurs most in region 1.

Fig 1.4 shows that most offenses seem to occur between 3 p.m. and 12 a.m. The occurrence rate differs by time across offense types.

Reference

1. Basic Analysis and Graphing, Second edition. Cary, NC: SAS Institute Inc.

Acknowledgement

We thank Dr. Goutam Chakraborty for his guidance on this project.