Crime Trends Analysis

Krithika Balan
CSCI 5502
Department of Computer Science
University of Colorado, Boulder
Krithika.Balan@colorado.edu

Keerthi Chikalbettu Pai
CSCI 5502
Department of Computer Science
University of Colorado, Boulder
Keerthi.ChikalbettuPai@colorado.edu

Pallavi Madasu
CSCI 5502
Department of Computer Science
University of Colorado, Boulder
Pallavi.Madasu@colorado.edu

1. INTRODUCTION

In this paper, we examine the impact of external factors that have affected crime rates across various cities in the last twenty years. We map crime data with outside influences such as weather, media violence from movies and series and other events such as natural calamities, economic crises and terrorist attacks across the country. We analyze the change in crime rates on the days immediately following the occurrence of any of these influences. This study examines whether certain events contribute to the increase or decrease in crime rates which allows society to be prepared accordingly. The results from this analysis could also help the police force to be better equipped to handle the most prevalent crime in any given neighborhood. Also, people moving into new cities can pick out safe neighborhoods to live in.

2. MOTIVATION

The number of shootings in the USA increased from 255 in 2013 to 331 in 2015[1]. There has been a steady increase in the number of shootings in the country over the years which illustrates that there might be a trigger causing the increase. This steady increase also applies to crimes other than shootings. Our study will examine the relationship (if any), between the sentiments and psychology of the people and the external triggers that lead to the increase or decrease of the crime rate at different time periods.

3. LITERATURE SURVEY

There are many existing studies analyzing how the crime trends have been influenced, individually by each of the external factors such as natural calamities [2][3], movies [4], weather [5][6], National Football League [7] and economy [8][9].

3.1 Impact of natural calamities:

Zahran, Shelley, Peek and Brody found that natural calamities in Florida tend to increase the number of domestic violence crimes in the state, but decreases the rates of property and violent crimes.

Roy analyzed the crime rates in certain parts of India and observed an increase in the crime rates, particularly property crime, following moderate to big disasters.

3.2 Impact of movies:

Dahl and Vigna conducted laboratory experiments in psychology to identify the relationship between movie violence and violent crime rates.

They observed, surprisingly, that the rate of violent crimes decreased on days when there were large number of viewers for violent movies.

3.3 Impact of weather:

Cohn, in her paper, analyzed the different researches that have been conducted to correlate crime and weather. She describes the effect of various weather conditions such as temperature, precipitation and humidity on different types of crime.

Zanten has compared crime reports available from 2001 with weather data for each day and found a pattern in the crime trends in accordance with temperature. This comparison was done for the city of Chicago.

3.4 Impact of economy:

A report prepared by United Nations Office on Drugs and Crime details the crime types in twelve different countries which are the most affected by economic factors. The report also lists the economic indicators which predict the change in these crime rates.

Andrienko studied the change in crime rates with the economic changes in Russia. He observed that the number of violent crimes increased with the increase in unemployment rate, but with a stable economy, there were more property crimes than violent ones.

4. PROPOSED WORK

There have been many studies focusing on a particular factor that affects crime. These studies have also been focused to a single state of the United States or have been conducted on other countries. Our project aims to consolidate these findings into a single paper which covers various factors that affect crime across various cities in the United States. We propose to find the patterns in crime rates with respect to a number of external factors as listed previously.

5. MILESTONES

We aim to answer the following questions in the course of the project:

1. How have major events such as elections and sports events affected crime rate over the years?
2. How has crime evolved through the years in a given neighborhood?
3. Is there a pattern in the evolution of crime over the years?
4. How have terrorist attacks affected crime rates?

Conference'10, Month 1–2, 2010, City, State, Country.

5. How have the crimes rates changed immediately following natural calamities?
6. Is there an increase in drug related crime after the release of drug related movies and series like Breaking Bad?
7. Is there a change in crime rate in accordance with the weather? Is there a change in the number of violent crimes with the release of movies involving violence?

6. DATASETS

We obtained the crime data for eleven different cities from the US City Open Data Census website. Historical weather data for these cities was downloaded from the Weather Underground[21] website.

Among all available data, the factors for choosing datasets were:

1. Data must be complete and freely available.
2. Data must be available for at least five years.
3. Cities are spread across the US.

Based on these criteria, we chose the following datasets for our analysis:

- Atlanta Police Department Open Data [10]
- Baltimore Police Department Open Data [11]
- Baton Rouge Crime Incidents [12]
- City of Boston – Crime Incident Reports [13]
- City of Chicago – Crimes Data Portal [14]
- Cincinnati Police Crime Incident Data [15]
- City and County of Denver Crime [16]
- Los Angeles – Current Crime Data [17]
- Louisville, Kentucky – Crime Data [18]
- City of Philadelphia Crime Incidents [19]
- San Francisco Police Department Incidents [20]

7. PEER REVIEW

In the session we discussed about various project topics, what each team expects to learn from the project and what results each team expects from the project. Furthermore, we discussed about where teams are finding their datasets, the nature of the data and if processing is needed. This stirred the session into discussing more about data cleaning, the various tools needed and how to handle missing data.

Few concerns that we thought were important for this paper were: How to combine multiple datasets? How to handle preprocessing for four datasets? Is there a need for specific tool for each dataset?

8. DATA PREPROCESSING

Since crime datasets were downloaded from different sources, one of the major issues was to standardize the types of crime. To do this, we identified the various types of crime in each of the datasets, classified them into a set of standard crime types. The set of standard crime types consisted of Assault, Larceny, Sex Offenses, Drugs/Alcohol Violations, Auto Theft, Weapons Trespass, Robbery, Burglary, Kidnapping, Fraud, Homicide, Arson, Gambling, Rape, Hooliganism and Traffic Violations. We then mapped the crime type in each of the datasets to one of these standard crime types. Some datasets also contained irrelevant records which correspond to events in the police beat, but not to actual crime events. We removed such records from the datasets as they would not be of use in our analysis.

Another issue we encountered was the format of Date and Time fields in the datasets. Since our analysis is based only on the dates of crimes, we extracted the dates from Date-Time fields for the datasets which do not have separate date fields.

While analyzing the crime patterns in various neighborhoods, we decided to ignore those records which do not contain address information. We also integrated the weather data for all cities considered with the crime data for each city in order to analyze the impact of weather on crime rates.

9. CLASSIFICATION AND PREDICTION

9.1 Correlation Analysis:

We used correlation coefficient to determine the relation between various weather parameters like temperature, dew point, humidity and so on. Following are the results we obtained for Baltimore:

Most Positively Correlated Attributes with crime rate: Temperature (0.55), Dew Point (0.51), Month in which the crime occurred (0.28) and Humidity (0.17)

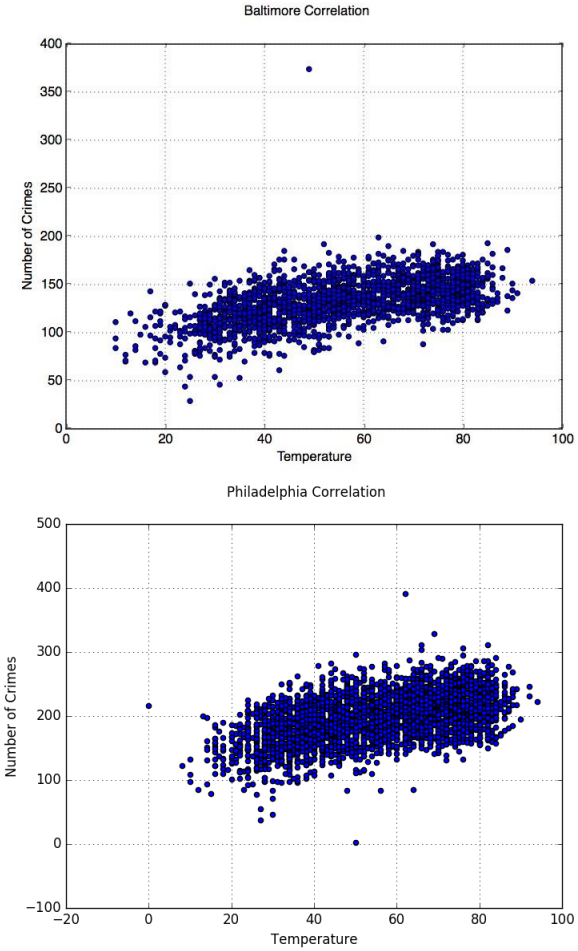


Fig. 1: Correlation between number of crimes and temperature in Baltimore and Philadelphia.

Most Negatively Correlated Attributes with crime rate: Wind Speed (-0.20), Sea Level (-0.23)

We obtained similar results for the rest of the cities too. So, these attributes are chosen as parameters to train the classifiers.

9.2 Decision Tree:

We implemented Decision Tree classifier to predict the crime type for a given set of input parameters.

Input – Temperature, Dew Point, Month of crime, Humidity

Output – One of the 31 crime types standardized as part of data preprocessing

The classifier is trained on 80% of the data by random split and tested on the remaining 20% (5- fold cross validation) and obtained an accuracy of 29%. After analysis, we found that the accuracy is less because of huge number of output labels, the classifier was not able to predict the occurrence of a particular crime type more than 29%. So, we went a step further to map closely related crimes like Burglary, various types of robbery and larceny, Auto theft to Theft, Vandalism, Disturbing the peace, Trespass, Vagrancy to Hooliganism and so on, thus reducing the unique crime types to 15 types. With the modified crime types, Decision Tree Classifier accuracy ranged from 69% – 92% for all the cities.

Below is the visualization of the decision tree. This can be also viewed here.

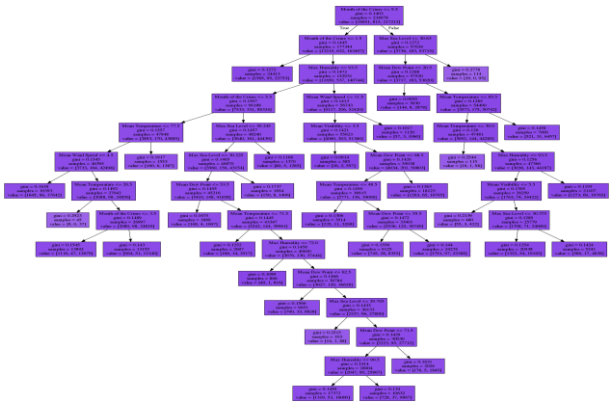


Fig. 2: Decision Tree classification for Baltimore.

9.3 Naïve Bayes:

Naïve Bayes classifier is implemented to predict the relative probability of the occurrence of various crime types given a set of input parameters. Accuracy ranged from 69% -92%. On an average we found that the probability of the occurrence of Theft is more compared to other crime types.

9.4 Random Forest:

We implemented Random Forest Classifier to predict the type of crime based on the input parameters and observed that the accuracy of the classifier has increased by 2 – 3% as compared to Decision Tree classifier. The reason for this is Random Forest is an ensemble classifier that constructs multiple decision trees and gives the mean prediction of these individual trees and controls the overfitting of the training data, hence increasing the accuracy.

10. KEY RESULTS

We started the analysis of the crime data by plotting the number of crimes against the date. We observed that the number of crimes is lesser in the beginning of the year. It gradually increases and then, reduces again for the latter part of the year. This variation can be seen in the graphs in Fig 3.

Additionally, in Baltimore, we can see a steep increase in the number of crimes in May, 2015. This could be an effect of the large number of shootings that occurred in Baltimore in the month of May, 2015. Incidentally, Baltimore had the most number of violent crimes in May 2015, when compared to the last 40 years[22].

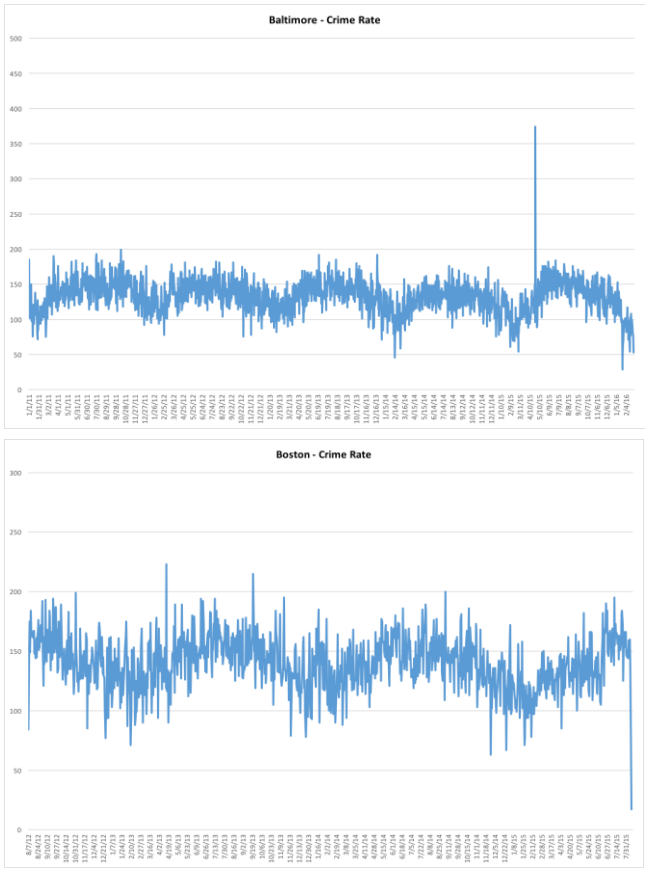
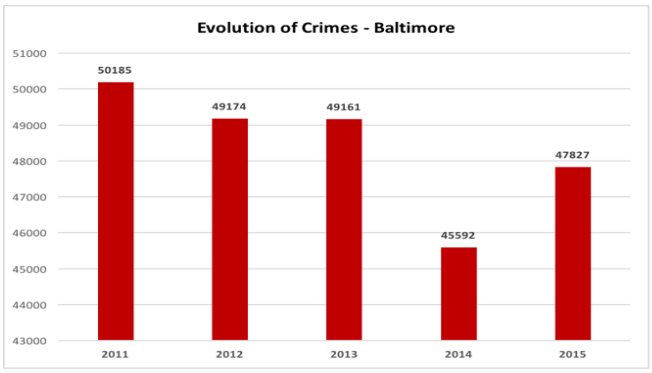
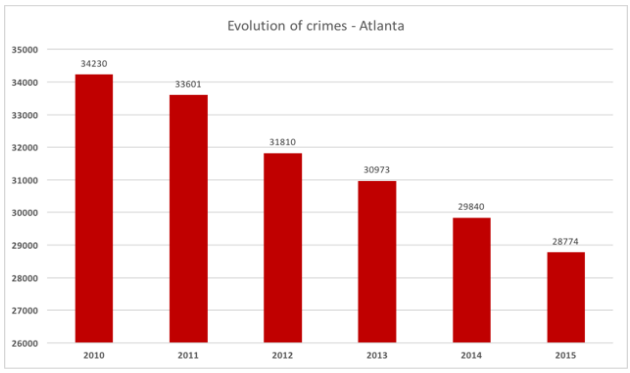


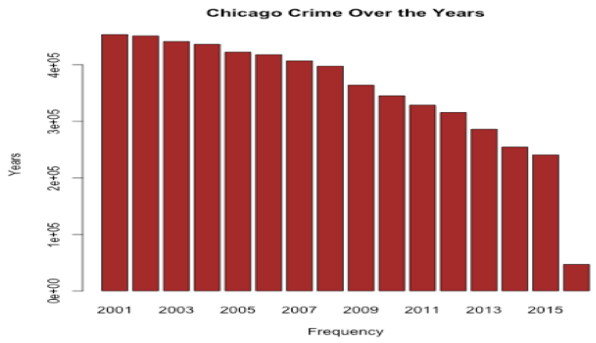
Fig. 3: Crime rates over the years in Baltimore and Boston.

Next, we analyzed the crime trends over the years in various cities. For this, we plotted the number of crimes for every year that we had data for a particular city. In general, we could see that the number of crimes have decreased in every city except Denver wherein we see an increase in the last three years. The increase can be attributed to an increase in population of Denver over the years, or maybe due to the legalization of marijuana in 2014. These trends can be seen in Fig 4.

In the next step, we plotted the number of crimes per month for different years of each city as seen in Fig 5. A common pattern that we observed is that the number of crimes dipped in February, increased in summer, peaked in fall, highest in August, and gradually decreased in winter.







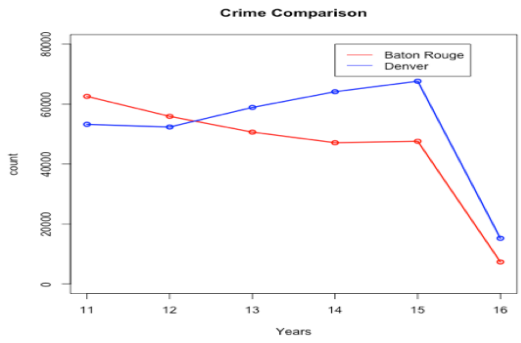
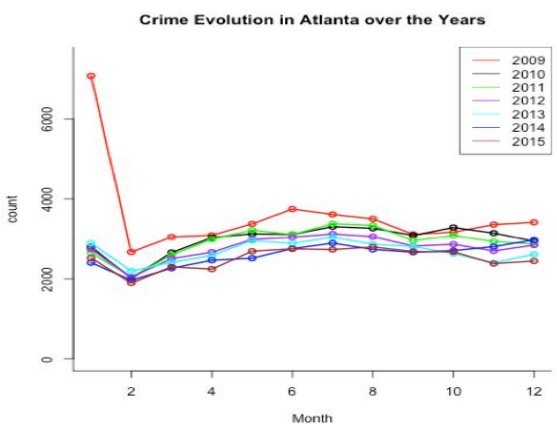
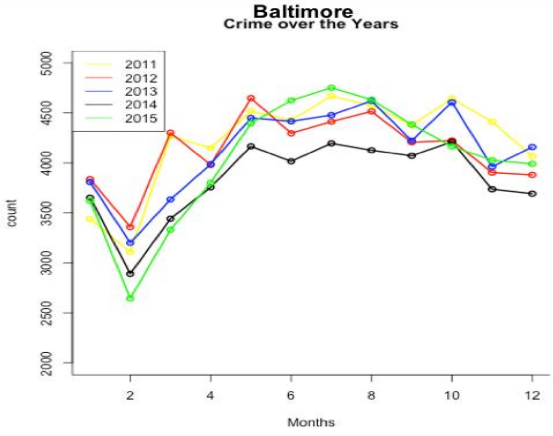
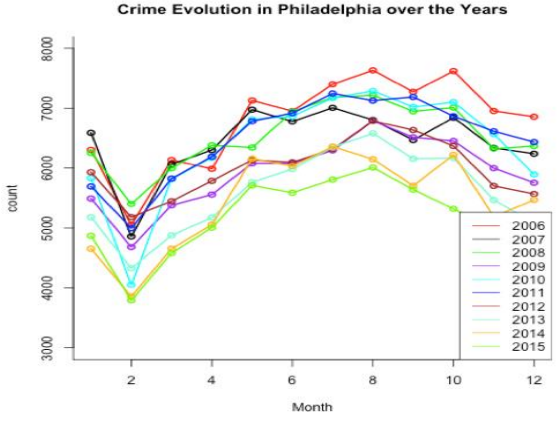


Fig 4. Change in number of crimes over the years in Baltimore, Atlanta, Chicago, Baton Rouge and Denver







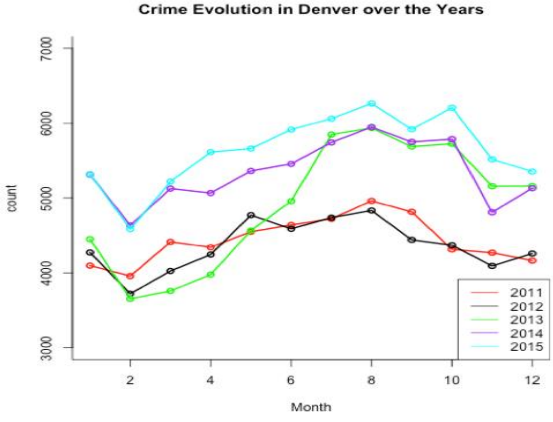


Fig 5. Number of crimes per month over the years in Atlanta, Baltimore, Philadelphia and Denver

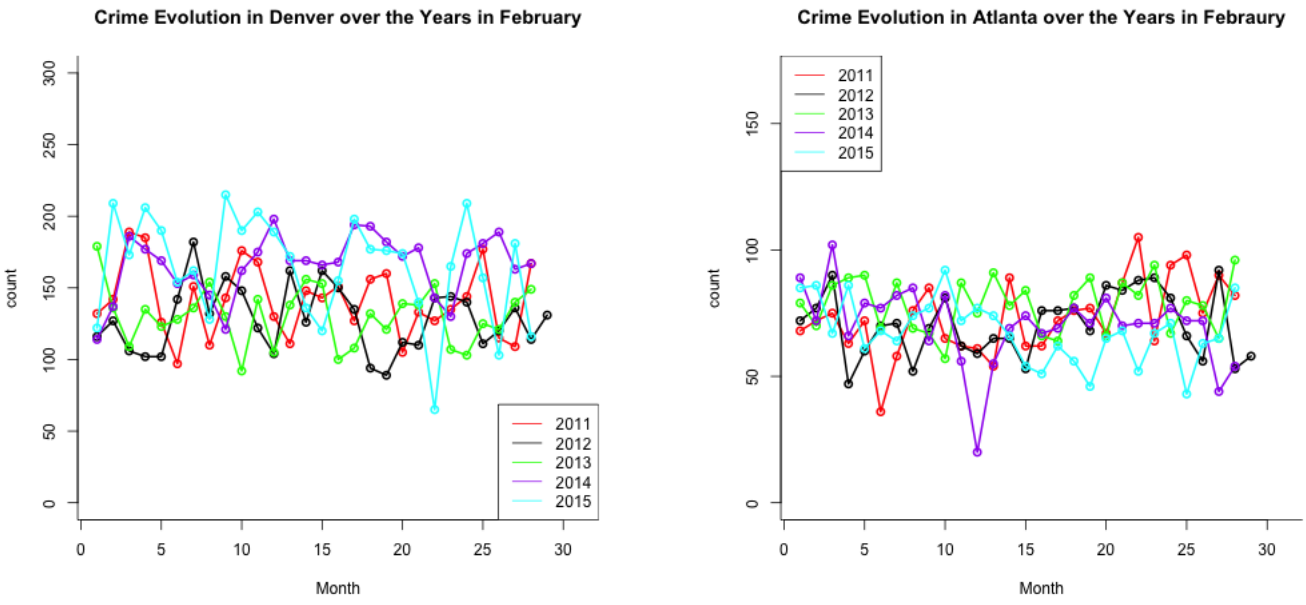


Fig. 6: Crimes in February over the years in Denver and Atlanta

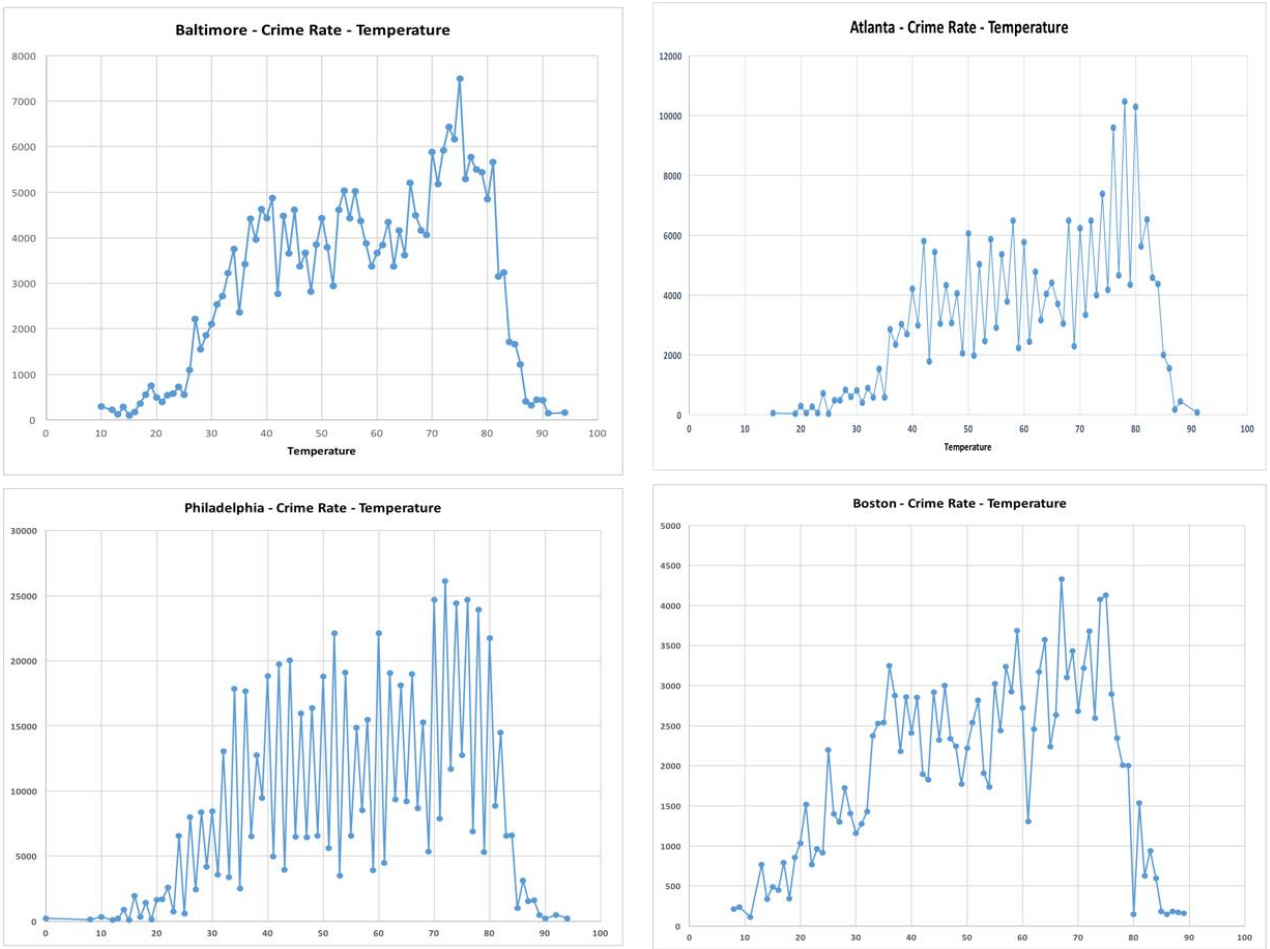


Fig 7 : Crime Rate vs Temperature – Baltimore, Atlanta, Philadelphia & Boston

Fig 8 : Crime Rate vs Dew Point– Baltimore, Atlanta, Philadelphia & Boston

To analyze this pattern further, we mapped the number of crimes in February for all the years across various cities. Fig 6 shows these plots for Denver and Atlanta. Few interesting observations we could see is that the number of crimes generally decreases slightly around Valentine's Day and on the days of the Super Bowl.

In the past few years, the Super Bowl has taken place on the following days: Feb 1 2009, Feb 7 2010, Feb 6 2011, Feb 5 2012, Feb 3 2013, Feb 2 2014 and Feb 1 2015. We can see in Fig 6 that these dates show a definite decrease in crime in both Denver and Atlanta. The graph also shows a relatively lesser number of crimes on and around Valentine's Day.

From the plot of crimes per month in Fig 5, we could see that the most number of crimes usually occur in August. To analyze this trend, we plotted the variation of number of crimes with weather in various cities. Fig 7 shows the variation of crime with temperature in Atlanta, Baltimore, Philadelphia and Boston and Fig 8 shows the variation for the same cities with dew point. We can see that up to a certain point, the number of crimes increases with increase in temperature and dew point. The fact that the number of crimes reduces after a specific high temperature is reached can be attributed to the fact that there might be fewer days with such high temperatures. In general, however the number of crimes seems to increase steadily with temperature.

But the above result does not hold good for individual crime types. For example, auto theft in Philadelphia does not always follow the pattern of increase in August and decrease by December. Also, the number of homicides per month varies randomly through the year. These results can be seen in Fig 9.

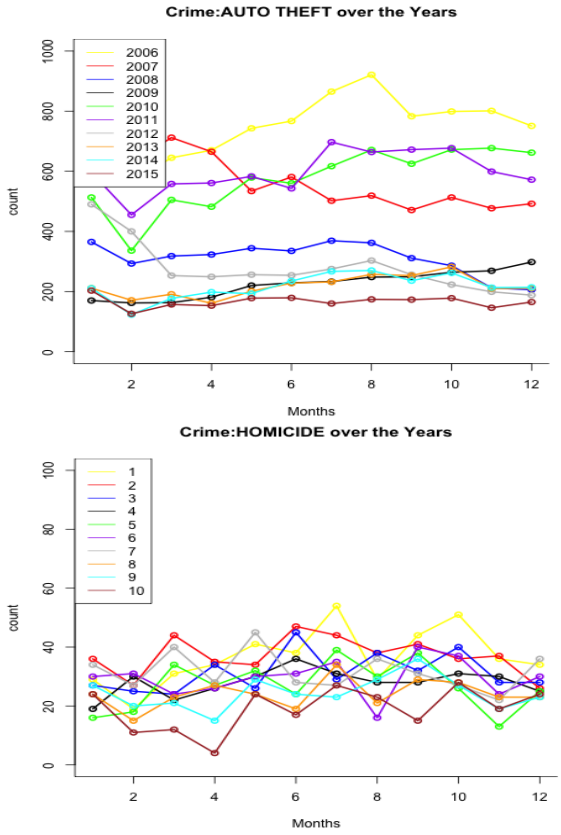


Fig. 9: Auto-Thefts and Homicides over the years in Philadelphia

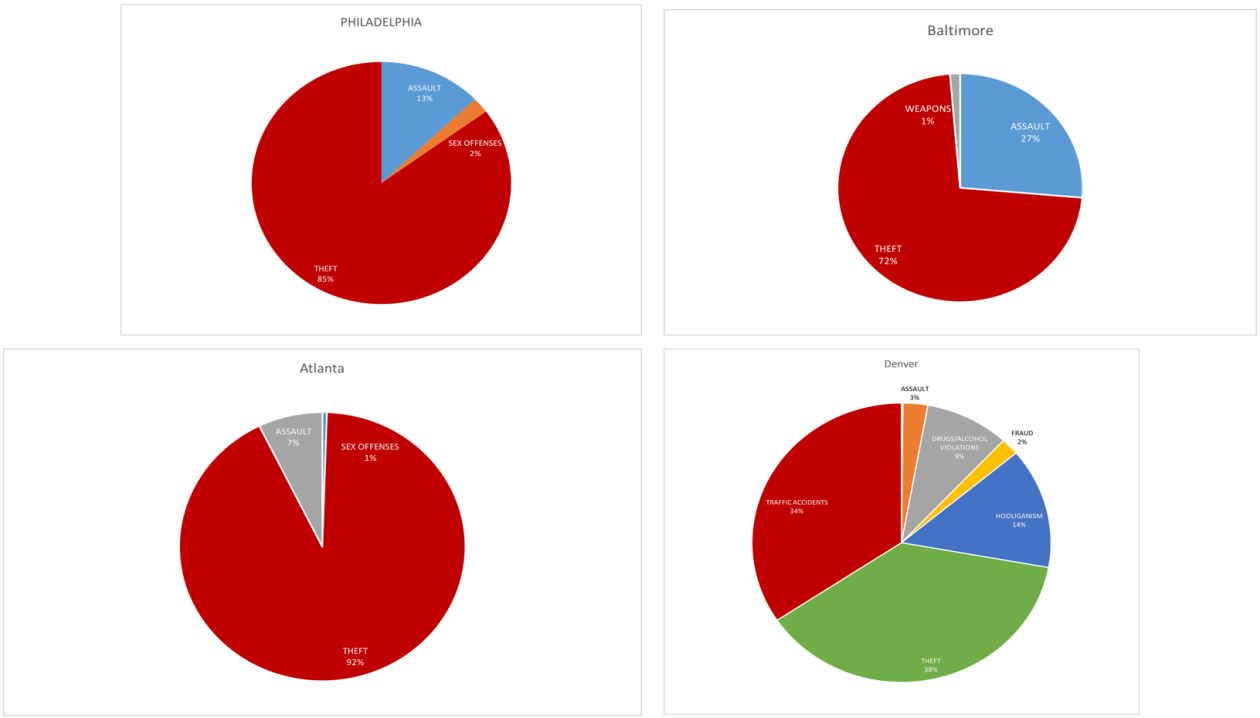


Fig 10 : Variations of Crime in Philadelphia, Baltimore, Atlanta & Denver

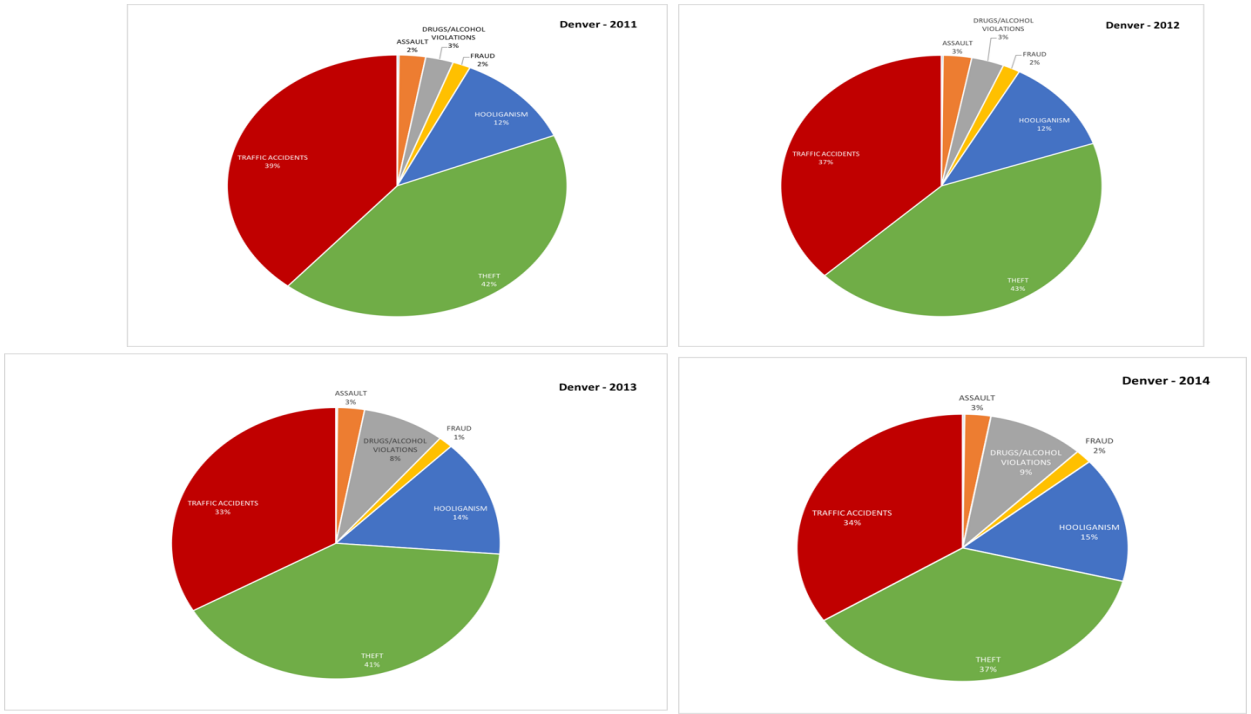


Fig 11 : Percentage of various types of crime in Denver from 2011-14

As a next step, we plotted the distribution of the different types of crimes in various cities. Fig 10 shows these distributions for Philadelphia, Baltimore, Atlanta and Denver in 2015. We then analyzed how various types of crime evolved over the years across cities. An interesting find during this analysis was a steep increase in drug and alcohol related activities in Denver since 2012 which is seen in Fig.11. This could possibly be attributed to the fact that the Denver Police Department changed the way the crime stats are reported in 2013. Furthermore, Amendment 64 was passed in 2012 to legalize the use of marijuana.

Our next aim was to check for any relation between the occurrence of certain events and the crimes. To perform this analysis, we picked the following events that could have had a potential effect on the crime numbers: Presidential election in

2012, terrorist attack on the twin towers in 2001, Boston Marathon bombing in 2013, Ferguson shooting incident in 2014, release of violent movies such as Django Unchained and The Expendables and natural calamities like Hurricane Sandy and the North American Cold Storm in February 2014.

While we could not see any major effect on crime numbers for most of these events, there was a definite impact of the Ferguson incident on crime numbers in Baltimore, and Atlanta, as can be seen in Fig 12. Fig 12 shows the plot of crime numbers in August over the years in Atlanta and the crime numbers in August and September 2014 for Baltimore. Both the cities have an African-American population over 50%. We can see a marked increase in the number of crimes on August 9, 2014 followed by a steep decline. This could be due to the protest immediately following the shooting of Michael Brown in Ferguson, Missouri.

A similar impact on crime was produced by the North American cold wave in February 2014. The sharp decrease in temperature in February caused a great decline in crime numbers in Atlanta as can be seen in the plot of Atlanta in Fig 6. This further iterates the effect of temperature on crime.

Finally, we analyzed the spread of crime in different parts of each city. This analysis helps identify the safer neighborhoods of each city. The heat maps in Fig 14 show the overall spread of crime in San Francisco, Atlanta, Denver and Chicago. We further analyzed the spread of each type of crime in the various cities. This can be seen in Fig 15 which shows the overall spread of crime in Philadelphia and the spread of auto thefts, robbery and assaults in Philadelphia.

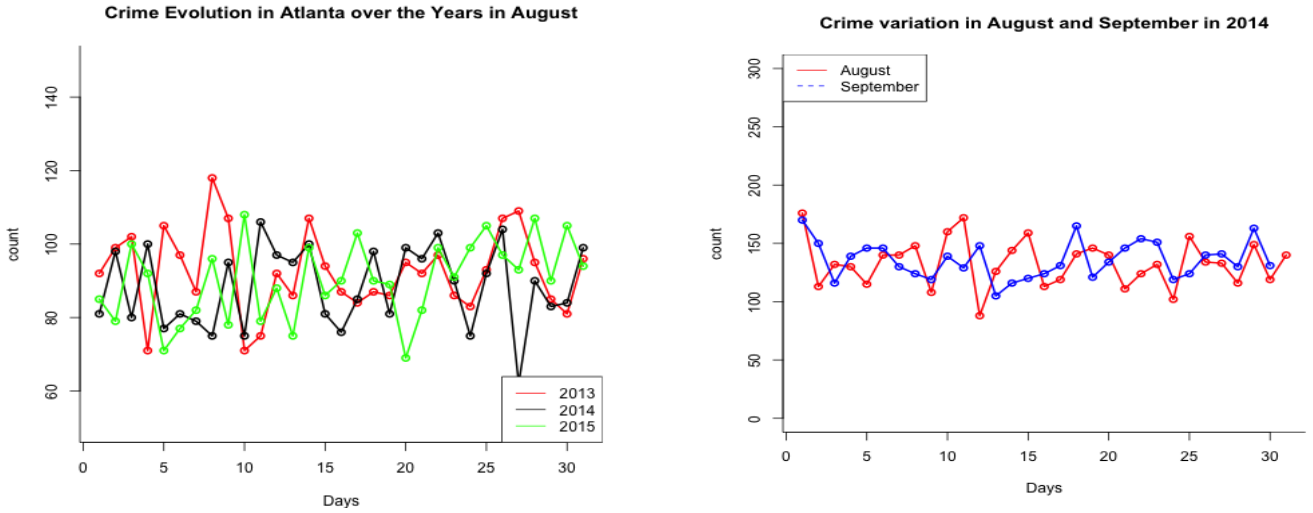


Fig. 12: Crimes in August and September over the years in Atlanta and Baltimore

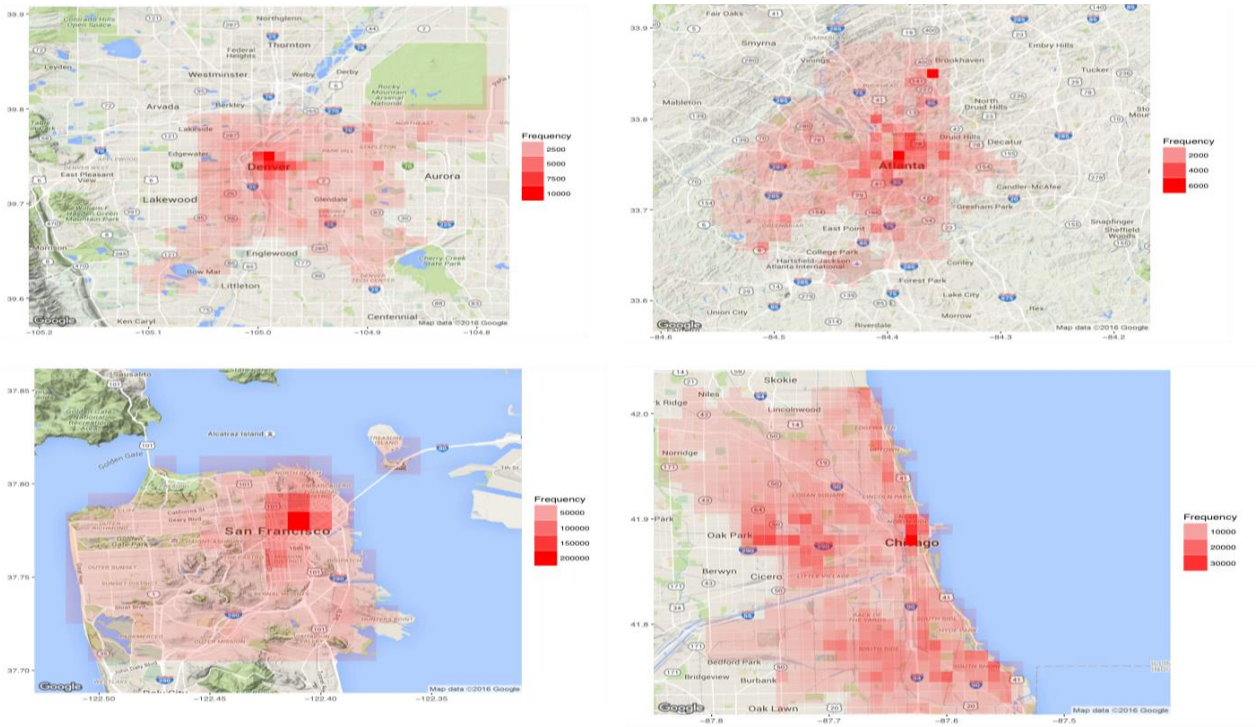


Fig. 13 Heat Maps showing overall Crime Hot Spots for Denver, Atlanta, San Francisco and Chicago

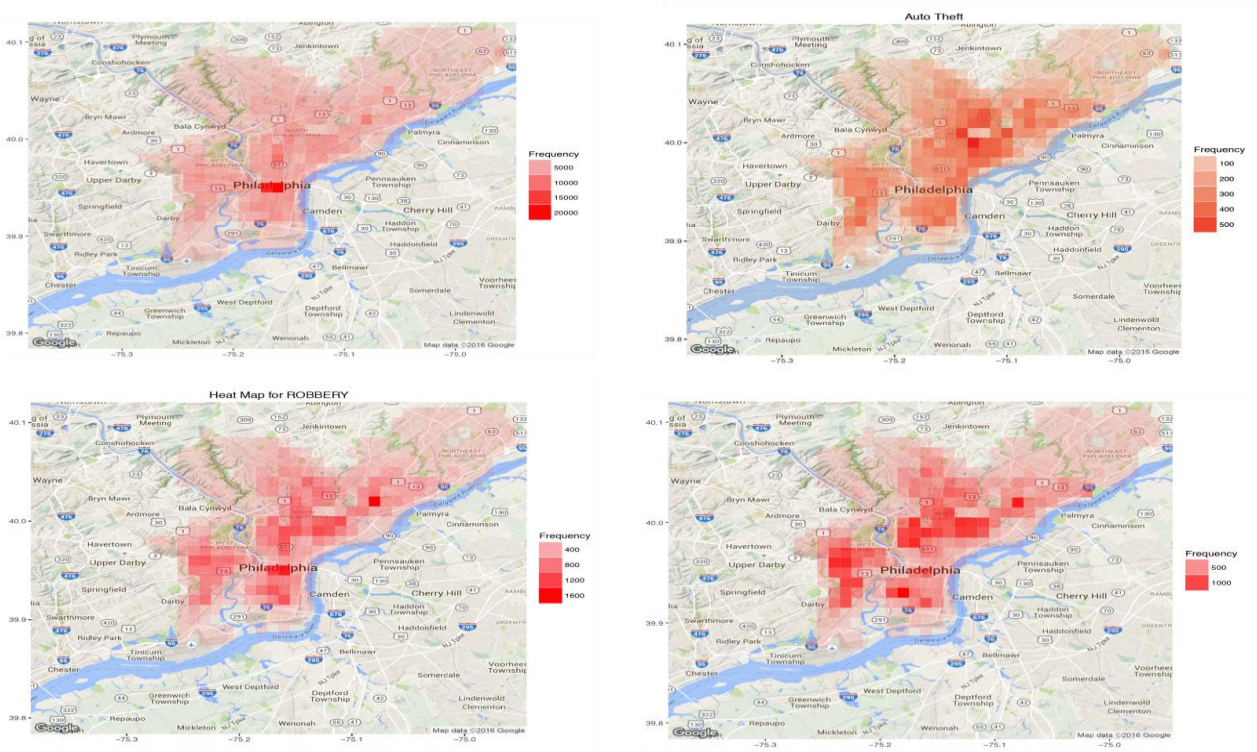


Fig. 14 Heat Maps showing overall crime spread in Philadelphia, Auto-Theft in Philadelphia, Robbery in Philadelphia, and Assault in Philadelphia.

11. EVALUATION CRITERIA

Accuracy: Decision-Tree and Naive Bayes classifiers are used to predict the crime type and the probability of occurrence of each crime type respectively with an accuracy ranging from 69% to 92% for various cities using 5-fold cross validation. Random Forest classifier outperforms Decision-Tree classifier by a 2-3% increase in accuracy depending on the city. This is because Random Forest is an ensemble classifier and the prediction is based on the mean values of multiple random decision-trees.

Robustness: Since we had to handle humungous datasets of various cities, our system is able to handle large data, is robust enough to integrate weather data and any additional crime datasets, and supports incremental changes to our current dataset.

Interpretability: For patterns observed in various cities, the crime trend can be interrelated to weather attributes such as temperature and dew point, and events like the Boston marathon bombing in 2012, Ferguson shooting in 2014. The results obtained from the correlation values between temperature and crime, and dew point and crime further strengthens our observations. The overall crime peaks in the month of August and is the least in the month of February for which the reasons are already highlighted above.

For our trend analysis, we manually looked at the patterns generated and determined if they made sense or not. Also, we mapped any anomalies found to various events, or/and attributed them to missing data or corrupted data.

12. CHALLENGES FACED

- Deciding on the various criteria to select the dataset was challenging because datasets for many cities were not open, reliable or complete. We also had to choose the cities based on their spread over the country.
- Crime is not reported in a standard way across cities and so, datasets for the various cities had to be made uniform manually, which turned out be a huge challenge. Firstly, we determined the crime types that had to be considered and the types that could be ignored. Secondly, we had to combine similar crime types in order to reduce 100 different crime types to 31. Finally, we had to identify the crime types that could potentially yield interesting patterns.
- While choosing the parameters to train the classifier, we had to find correlation between various weather attributes and the total number of crimes. In addition to this, we intuitively decided to find correlation with other parameters which may influence crime such as month of the crime, number of crimes that occurred in each month/year, and so on.
- Initially we decided with 31-stanrdard crime types on which we concentrated in this project. On feeding these crime types to the classifiers, the results obtain had an accuracy of 29% with 5-fold cross validation. We then reduced the standard crime types to 15 which drastically increased the accuracy of the classifier to 69% with 5-fold cross validation.

- Given that we were dealing with various cities having numerous crime types for a range of 5-15 years, finding consistent patterns and manually deciding if these patterns were interesting was a huge challenge.
- In order to find safe neighborhoods in a city, we divided the city into 10x10 grid based on the latitude and longitude coordinates of the crimes in the dataset for the city. This yielded 100 neighborhoods and we then calculated the number of crimes in each neighborhood. This was made difficult by the fact that there were many outliers and missing data in the datasets.

13. FUTURE WORK

We can extend our analysis on crime further in the following direction.

- By considering historic crime datasets extending up to (say) 40-50 years for various cities, we might find more compelling trends in crime having higher correlations.
- We considered few events and movies, based on the selected cities and duration of the data available, such as the Boston marathon bombing in 2013, Ferguson shooting in 2014, Breaking Bad, Blood Diamond, and so on. We can extend this analysis further with a historic crime dataset and find better correlation between these events/movies with crime.
- We can also consider the impact of economy on crime, and the variation of crime with the demographics of the city.
- In this project we concentrated on the main crime types like robbery, larceny. We can future extend this work by analyzing the trends in sub-crime types like robbery-pedestrian, robbery-commercial, robbery-residence, larceny-vehicle, larceny-nonvehicle.

14. REFERENCES

[1] *Gun Violence Archive*. N.p., n.d. Web. 25 Feb. 2016. <http://www.gunviolencearchive.org/>.

[2] Zahran, Sammy, Tara O'Connor Shelley, Lori Peek, and Samuel D. Brody. "Natural Disasters and Social Order: Modeling Crime Outcomes in Florida." *International Journal of Mass Emergencies and Disasters*. N.p., n.d. Web. 25 Feb. 2016.

[3] Roy, Susmita. "The Impact of Natural Disasters on Crime." (n.d.): n. pag. Web. 25 Feb. 2016.

[4] Dahl, Gordon, and Stefano DellaVigna. "Does Movie Violence Increase Violent Crime?" (n.d.): n. pag. Web. 25 Feb. 2016.

[5] Cohn, Ellen G. "WEATHER AND CRIME." *The British Journal of Criminology* 30.1 (1990): 51-64. Web. 25 Feb. 2016.

[6] Zanten, Eric Van. "Crime vs. Temperature." *Crime vs. The Weather*. N.p., n.d. Web. 25 Feb. 2016

.

[7] Peña, Alejandra Iraide. "An Evaluation of the Impact of the NFL's Super Bowl Event on the Host Cities' Crime Rates." *An Evaluation of the Impact of the NFL's Super Bowl Event on the Host Cities' Crime Rates*. N.p., n.d. Web. 25 Feb. 2016.

[8] Malby, Steven, and Philip Davis. "Monitoring the IMPACT OF ECONOMIC CRISIS ON CRIME." *UNITED NATIONS OFFICE ON DRUGS AND CRIME* (n.d.): n. pag. Web. 25 Feb. 2016.

[9] Andrienko, Yury. "Explaining Crime Growth in Russia during Transition: Economic and Criminometric Approach*." *Policy Documentation Center* (n.d.): n. pag. Web. 25 Feb. 2016.

[10] http://opendata.atlantapd.org/Crimedata/Default.aspx

[11] https://www.baltimorepolice.org/bpd-open-data

[12] https://data.brla.gov/Public-Safety/Baton-Rouge-Crime-Incidents/fabb-cnnu

[13] https://data.cityofboston.gov/Public-Safety/Crime-Incident-Reports/7cdf-6fgx

[14] https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2

[15] https://dev.socrata.com/foundry/data.cincinnati-oh.gov/gr2k-bw2d

[16] http://data.denvergov.org/dataset/city-and-county-of-denver-crime

[17] http://shq.lasdnews.net/CrimeStats/CAASS/desc.html

[18] http://portal.louisvilleky.gov/dataset/crimedataall-data

[19] https://www.opendataphilly.org/dataset/crime-incidents

[20] https://data.sfgov.org/Public-Safety/SFPD-Incidents-from-1-January-2003/tmnf-yvry

[21] https://www.ncdc.noaa.gov/data-access/quick-links

[22] http://www.usatoday.com/story/news/nation/2015/05/31/baltimore-homicides-worst-in-40-years/28284839

15. APPENDIX

A. Honor Code:

On my honor, as a University of Colorado at Boulder student, I have neither given nor received unauthorized assistance on this work.

B. Individual Contributions:

Data collection and data pre-processing was divided among all three team members based on the cities.

Krithika Balan:

1. Worked on crime evolution in a given neighborhood/city over month/year.
2. Designed a grid-based algorithm to divide a given city into neighborhoods based on latitudes and longitudes.
3. Programmed in python (libraries: pandas) to identify safe neighborhoods and distribution of crime in various neighborhoods.
4. Analyzed the effect of outliers and inconsistent data on these trends.

Keerthi Pai:

1. Worked on crime evolution in a given neighborhood/city over month/year.
2. Programed using R to plot various trends in crime for various parameters.
3. Analyzed the impact of events and movies on crime.
4. Plotted heat maps in R to visualize the crime hot spots in various cities.

Pallavi Madasu:

1. Worked on crime evolution in a given neighborhood/city over month/year.
2. Found interesting patterns by visualizing the trends in crime against weather attributes such as temperature, dew point, humidity and other crime influencing attributes such as month of the crime, number of crimes in each month and so on.
3. Programmed using python (libraries: pandas, scikit, matplotlib) to find correlations between weather attributes and various crime parameters.
4. Implemented the Decision-Tree, Naïve Bayes and Random Forest classifiers to predict the crime types and their probabilities for a given test data.