The relationship between weather and crime

Yun Gao
New York University
New York, USA
Yg1200@nyu.edu

Gulin Fu
New York University
New York, USA
gulinfu@nyu.edu

ChangYuan Li
New York University
New York, USA
Cl3721@nyu.edu

Abstract—**(Use a short version of your project proposal here.)**

An analysis of correlation between the weather and crime records in Chicago yielded strong evidence that weather has a positive effect on most types of crime. The analysis is based on weather data and crime data in Chicago from 2001 to 2015. Also, we identify the effect of the severe weather event on crime rate. To parse the big data of the crime record, Hadoop is used in this project. The result of crime rate in diverse weather conditions is presented the by charts.

Keywords—analytics

I. INTRODUCTION

The project focus on the correlation between the weather and crime records in Chicago. We combine weather data and crime data together to analyze the crime rate in different weather conditions. Also, we are interested about the crime rate of the extreme weather. Finally, we can present the crime rate in diverse weather conditions by charts.

1. Name of Data Source : Chicago data portal

Data Source Description:

This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days.

Data Collection Frequency: Static

Data Size: 2GB

Data Frequency: just once

2. Name of Data Source : NOAA

Data Source Description:

The weather data of Chicago from 2001-present Data Collection Frequency: Static

Data Size: 200MB

Data Frequency: just once

3. Name of Data Source : Severe weather

Data Source Description:

The start date and end date of the severe weather (Hall, flash flood, thunder etc.) of USA

Collection Frequency: Static

Data Size: 200MB

Data Frequency: just once

The typical user of our application would be police and security department. We can verify the goodness of the result based on the common sense and the related papers.

II. MOTIVATION

As New Yorkers, we all know that there has never been a day that people in New York can sleep without sirens and there has never been a day that no criminal is arrested in New York City. The public safety gradually become the serious problem for the metropolis including but not limited to New York City.

To solve this problem, it would be a potential solution by forecasting the crime rate every day and arranging the public safety resource by the various crime rate.

Benefiting from the Hadoop, now we are able to parse a huge number of crime records and the weather data, which means it is not that hard for us to figure out the relationship between the weather and crime. The department of public safety can arrange the police resource in a more effective way by referring the forecast.

III. RELATED WORK

There are many factors lead to "crime", and scientists are studying in this topic. For example, the accent is one of the reason make the police think someone is guilty. The authors summary the previous studies in the same topic, and using their conclusions to provide a hypothesis the accents may

influence the judgement of the guilty. The authors think a Brummie-accented suspect would elicit stronger attributions of guilt than a standard-accented suspect, and then, they did an experiment. The participants were asked to listen to a recorded conversation that was based on a transcript of an interview, also the authors made a lot of rules to eradicate the difference, such as different crime types, white collar or blue collar, and races. At last, the authors made the conclusion that nonstandard speakers are perceived as guiltier than standard speakers. [1]

Of course, weather is one of the most obvious factors may influence the crime, and many papers discuss their relationship. JAMES E PAKIAM explained the factors may lead to crime, including physical environment factors, socioeconomic environment, anthropological-biological factors and mental factors. Of course, the weather is one of the factors lead the crime, and the solar, temperature and humidity play crucial roles in the weather. Also, the paper review the history of people engaged in studying the relationship between weather and crime, the earliest scientific study to relate crime to the seasons was made by several scientists, they found in Germany, Aschaffenburg and Mannheim, several climate lead to the high crime ratio. At last, the author discusses the comfort zone for people, the people in this zone will feel good may lead to less crimes.[2]

This paper is the studying about the correlation between property and violent crime and temperature. Since the temperature can be affected by many factors such as rainfall and sunshine. The author also analyzed the possible correlation between these factors. To draw the final conclusion, the author calculated every Correlation Coefficients of the factors and found that only the temperature has a positive effect on the property and violent crime. [3]

Overall, due to the challenges of developing an econometric approach that can generate unbiased estimates of long-term relationship between weather and crime. The results of this paper suggest that climate change will have substantial effects on the prevalence of crime in the United States. Although this approach does not account for longer-term adaptation possibilities, I view it as the most credible way to provide useful information about the potential magnitude of the effects of climate change on future criminal activity in the United States. [4]

The paper discussed how climate change may be impacting crime rates. Based on analysis of 20 years of monthly data from St. Louis, Mo, it concludes that most major crime types are likely to be impacted by rising temperatures. It mainly use three tables to back up their conclusion: Table 1-Descriptive statistics, dependent and independent variables,Table 2-Time-series analysis, Table 3-Predicted impact for current and comparison studies. That is useful for us to learn how to lead to a conclusion with aggregate data. Differ from this

paper, we will use more attribute (time, temperature, day/night, weather type) to analysis.[5]

This paper did the research in two dimensions. The first dimension is study. They have 2 studies: a violent crime-weather relationship and a non violent crime-weather relationship. The second dimension, they made 2 predicates and then confirmed them both. In all, 2*2 = 4 relation model was provided to offer a quite convincing result. They use Figure 1. The difference between violent and nonviolent crime rates by quarter of the year, United States, 1971–1980 and Figure 2. Effects of yearly differences in temperature on violent and nonviolent crime rates, United States, 1971–1980 to back up their conclusion, simple figures though, were extracted from detailed math calculation. Multiple dimension study method is what can adopt for better confidence level.[6]

IV.   DESIGN
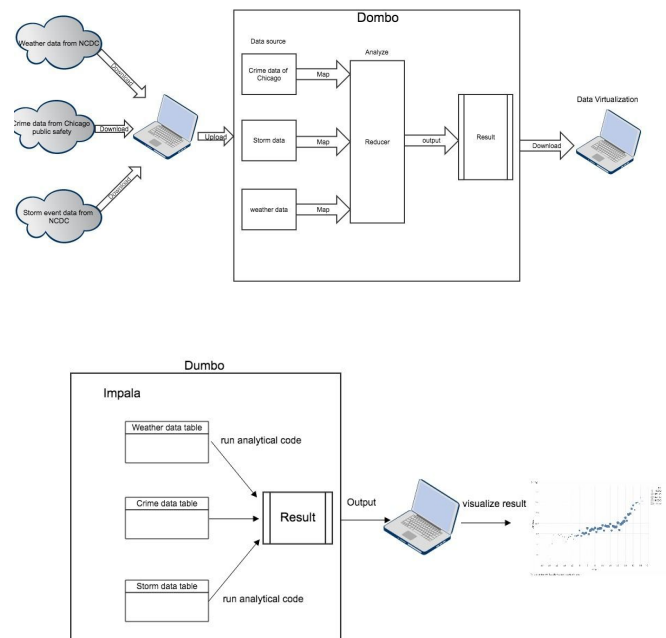
V.   RESULTS

The experiment data are mainly derived from three sources: the crime records from Chicago from 2001 to present, the hourly weather data in Chicago and the storm event data from National Centers for Environmental Information.

To acquire the essential information, the data shall be cleaned and profiled. Based on the experiment design, Hadoop mapreduce was used to clean and profile the data. To be more precise, the crimeid, crime time, crime types were extracted from the crime dataset, and the weather recorded time, location, wind speed, visibility distance, temperature, liquid precipitation depth dimension were extracted from the weather dataset. As for the severe weather records, startime, endtime and weather type were extracted to check the crime rate

between severe weather weather. After cleaning and profiling data, there were 6200661 crime records , 210030 weather records and 8227 severe weather records.

Since the amount of the records is large and our laptop can no longer parse the data, all the data was uploaded to NYU Dumbo and stored in several tables of impala. The most significant part of the experiment is to combine all the data together by time. Since there is no exactly same time of these records, we should join these tables by the nearest time of each records. Honestly, this was the most challenging problems for us since the time of each record was formated differently and the worse new is that one of the tables has 6200661 crime records. Even using NYU Dumbo, it took more than 8 hours to finish join operation. So the new approach was applied to analyze, first, the crime types were divided into five new types, including the sex crime, violent crime, property crime, organized crime and other crime, and the new table was created to store it. Secondly, instead of using the whole table to merge, five factors were considered separately, every time one factor was used in analyzing. Thirdly, the average factor data recording in every unique hour was calculated, then merge it into new crime table. At last, using the table to calculated the average crime rate in different weather conditions.

Further, to figure out how the crime rates would change during severe weather, 3 step of work was done. First counted the total crime numbers in each crime type, calculated their crime rate respectively. Then counted the total number of crimes in each crime type during different severe weather and the total hours for each severe weather in 15 years, combined them to get crime rate during severe weathers. Last compared the severe-weather crime rate and average crime rate to make our conclude.

Finally this approach is able to reveal the crime distribution on various weather conditions.

VI. FUTURE WORK

In the experiment, factors was considered separately, in the future, more than one factors shall be used to analyze concurrently. Meanwhile, the Chicago is a big city in United State, so one thing should mind, does the conclusion also applied to the rural areas? More data should be collected including the cities and the rural areas to prove the results.
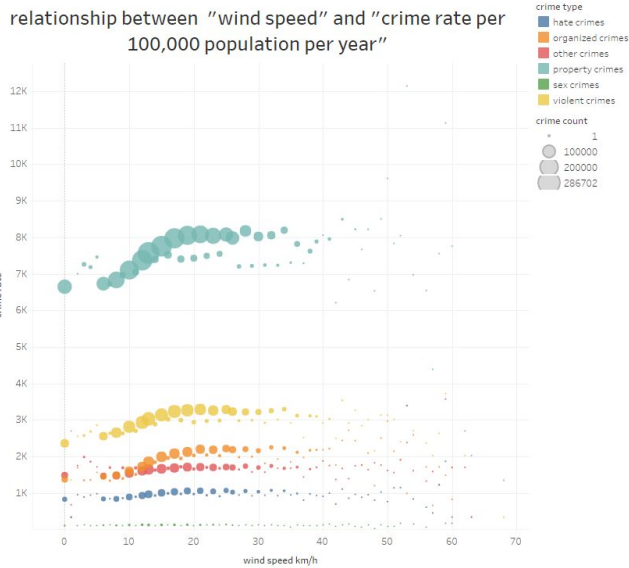
VII. CONCLUSION

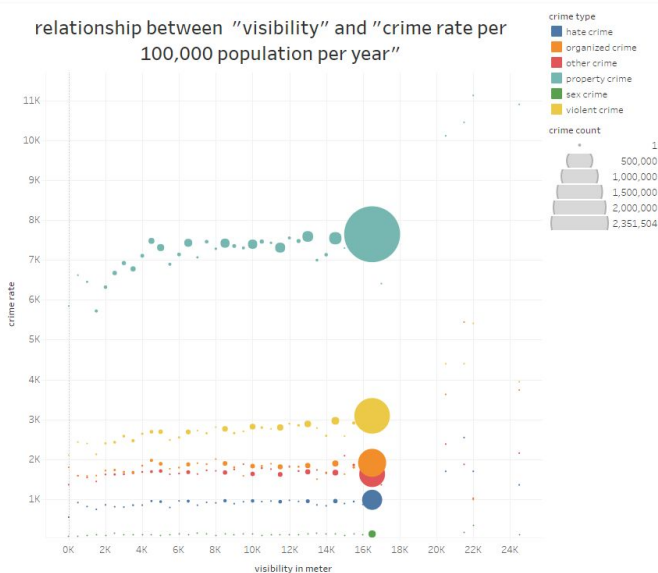1.Crime rate of all type increase while temperature raise.
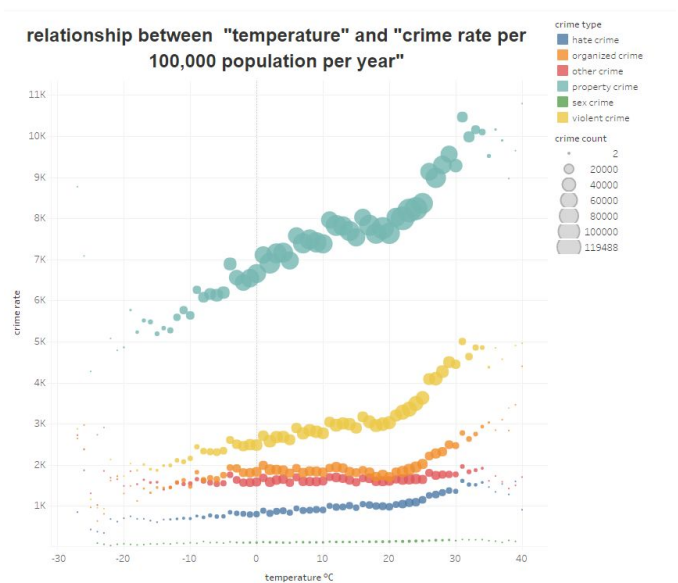
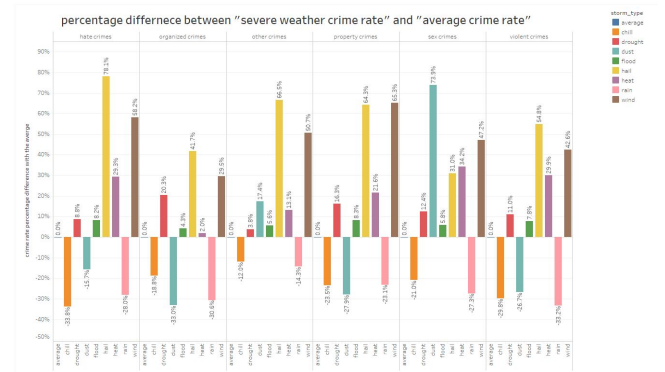2.Crime rate of all type decrease while rainfall precipitation raise.
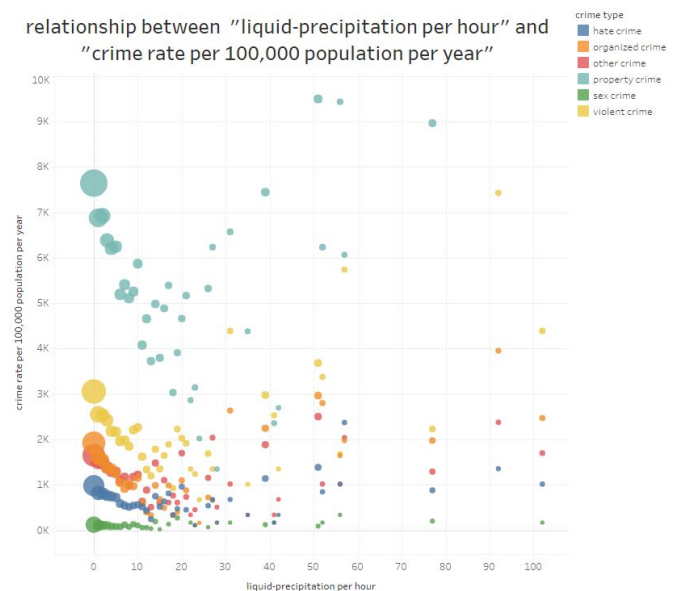
3.Crime rate of all type increase while wind speed raise amon range 0-15 km/h.After that, crime type remain about the same.

relationship between "wind speed" and "crime rate per 100,000 population per year"

percentage differnece between "severe weather crime rate" and "average crime rate"

4.Crime rate is almost non-correlation with visibility.

relationship between "visibility" and "crime rate per 100,000 population per year"

5.During hail, heat, drought, flood and heavy wind severe weather, crime rate tend to increase. During chill, dust, heavy rain severe weather, crime rate tend to decrease. But sex crime rate during dust weather increase dramatically instead of decreasing.

REFERENCES

[1] Dixon, John A., Berenice Mahoney, and Roger Cocks. "Accents of guilt? Effects of regional accent, race, and crime type on attributions of guilt." Journal of Language and Social Psychology 21.2 (2002): 162-168.

[2] Pakiam, J. E. 1981. "The effects of the weather on crime.". International Journal of Comparative and Applied Criminal Justice, 5(2): 179–186.

[3] Ranson M. Crime, weather, and climate change[J]. Journal of environmental economics and management, 2014, 67(3): 274-302.

[4] Field S. The effect of temperature on crime[J]. British Journal of criminology, 1992, 32(3): 340-351.

[5] Mares, Dennis. "Climate change and crime: monthly temperature and precipitation anomalies and crime rates in St. Louis, MO 1990–2009." Crime, law and social change 59.2 (2013): 185-208.

[6] Anderson, Craig A. "Temperature and aggression: effects on quarterly, yearly, and city rates of violent and nonviolent crime." Journal of personality and social psychology 52.6 (1987): 1161.