Visual Analysis of New York City Crime and Weather Data

Gang Zhao

Abstract—According to studies of the relationship between crime and weather around the world, weather does have a influence on crime activity. However, there's no systemic study of New York City on this topic since Dr. Edwin Grant Dexter's work in 1900s. To testify and improve former studies, this paper carried out a specfic case study of New York City on the relationship of crime activities and weather conditions based on weekly reported crime data from New York City Police Departemnt (NYPD) and weather data from Weather Underground in the year of 2012. The temporal and spatial distribution patterns are investigated through the city, borough and police precinct scales with Tableau. With the help of linear regression models, main weather factors that affect different crime types and prediction equations on different scales are identified. The findings and limitations of this paper, and potiental improvements in the future are also discussed to benefit any possible further studies.

Index Terms—Visual analysis, New York City, crime and weather, temporal and spatial distribution, linear regression models.

INTRODUCTION

In nowadays, more than half of all people live in cities, and over one million people will move to cities every week from now to 2050. The cities and urbanization have become progressively significant when we faced with various problems. In terms of scaling law in urban science, larger cities not only have more people, more economic activities, more innovations, more roads and more buildings but also more crimes. Understanding the crime patterns in big cities and thus improve the quality of life in cites is becoming increasingly important. New York City, which is the most popular city in the United States or even in the world, provides a great sample for conduct such studies.

The studies in criminology and criminological theory around the world strongly suggest that there's a relationship between crime and weather. These researches have successfully provided some people and city agencies with knowledge on how to better prepare themselves during certain weather conditions that might produce influences on crime activities. But whether these findings are applicable in New York City is not still unknown due to a lack of related data and studies. With more and more data and tools available through the Internet, these studies are no more unreachable.

The purpose of this paper is to utilise the data of New York City to testify and further investigate the relationship between crime and weather, and also to extend the application possibilities of visual tools in such researches.

1 RELATED WORK

As early as 1900s, Dr. Edwin Grant Dexter, who was an educator[1], had talked about the influences of weather on crime in his book *Weather Influence*[2]. In the chapter of *Crime and The Weather* from page 141 to 165, he combined the total number of arrests for "Assault and battery" and weather data which include temperature, barometer, humidity and wind in New York City from the year of 1891 to 1897 to investigate the relationship between them. As one of the first systemic studies in this area, Dexter made several far-reaching findings: except for the extremely high temperatures, the number of assaults goes up as the heat goes up; less assaults with higher barometers; a huge drop in number of assaults when the humidity reaches 0.9; more crimes in mild winds that between 150 and 200 miles per day. The ways Dexter used to analyse these data are mainly based on simple line charts, his personal experience and ideas from relative researches. Dexter's study with New York City crime and weather data carried out 100 years ago successfully provides a sensory and empirical analysis method and a contrast sample to this paper, since few similar research obtainable in New York City afterwards.

In the article *Weather and crime* from the winter published *British Journal of Criminology* in 1990[3], Ellen G Cohn combed former researches on crime and weather comprehensively and further discussed the future possibility to build a solid theoretical basis for crime and weather studies based on routine activities theory or rational choice theory. Through gathering and investigating all the previous researches covering numerous weather conditions and different kinds of crimes, Cohn identified some significant efforts made to advance the knowledge in this area: a positive relationship is found between the number of robberies and the number of days with a temperature less or equal to 32ºF; barometric pressure has no directly influence on the rates of assaults or homicide, but the barometric fluctuations might be caused by other changes in temperature or other weather variables; there's a slight negative correlation ($r = -0.11$) between humidity and the number of assaults. These findings would be carefully examined in this paper through visual analysis of crime and weather data of New York City to develop a deeper understanding of crime patterns in New York City, and thus make any possible contribution to the development of crime and weather study.

Both linear and curvilinear models have been used to carry out the quantitative studies on the relationship between crime and weather[4]. Since the curvilinear effect usually turned to be weak between the number of crimes and weather conditions, and curvilinear model also has over fitting risks, more researchers choose to use linear models. Butke and Sheridan utilized linear models to explore the relationship between apparent temperature and aggressive crime counts in their paper *An analysis of the relationship between weather and aggressive crime in Cleveland, Ohio* that published in April 2010[5]. Their analyses covered citywide level, sub city level and different time scales using daily temperature and crime data. According to their research, there's no obviously change in crime spatial patterns when temperature changes, and the number of aggressive crimes usually increases linearly as the temperature goes up. This analysis approach inspires the author with the idea and method of how to investigate the relationship between crime and weather in New York City in different spatial scales including city, borough and police precincts with visual methods and linear models.

In this paper, former noteworthy findings in weather and crime studies would be tested by the visual analysis of the data from New York City in city, borough and precinct scales, linear regression models are chosen to create different predication equations in different scales and compared to the observed crime number.

2 DATA AND SOLUTION

The data needed in this paper are all collected from the Internet. The data resources and the contents of analysis are described.

• *Gang Zhao is with New York University, Center for Urban Science and Progress. E-mail: gang.zhao@nyu.edu.*
•

2.1 Data

New York City crime data are downloaded from PediacitiesNYC[6] which include the weekly reported number of six types of crimes like murder, rape, robbery, assault, burglary and grand larceny (also a subclass grand larceny auto), from New York City Police Department (NYPD) based on police precinct in 2012, and one thing to note here is that the data of the week 4, 5, 9, 15, 23, 32, 35, 39, 46, 50 are missing.

The daily weather data are available from Weather Underground[7], and weather variables like temperature, humidity, sea level press and visibility and wind are included. To combine the weather data with weekly reported crime data, the different precincts from the same borough use the same weekly averaged weather data from the nearest weather station, for example, the precincts from Manhattan and Brooklyn use the weather data from Central Park station, the precincts from Staten Island use the data from Newark station, the ones in Queens and Bronx use the data from New York JFK station and New York station.

2.2 Solution

To achieve a better understanding of the crime patterns in New York City, three scales are investigated in this paper as following:

City scale: compare total number of five types of crime in the city; find the temporal distribution patterns of total number of crimes and each crime in the city; compare the influences of different weather variables on different types of crimes in the city; build predication model on city scale.

Borough scale: identify the spatial distribution of five types of crime in the boroughs, build predication models on borough scale.

Precinct scale: identify the spatial distribution of five types of crime in the precincts; find the spatial distribution patterns of total number in each precinct; build predication models on precinct scale.

The visual analyses are mainly carried out in Tableau 8.0, and the linear regression models are built in SPSS 18.

3 Results

The weather's influences on crime activities are separately discussed in city, borough and precinct scales as following:

3.1 City scale

The number of crimes of each type in 2012 are assigned into linear dimensions according to different weather conditions.

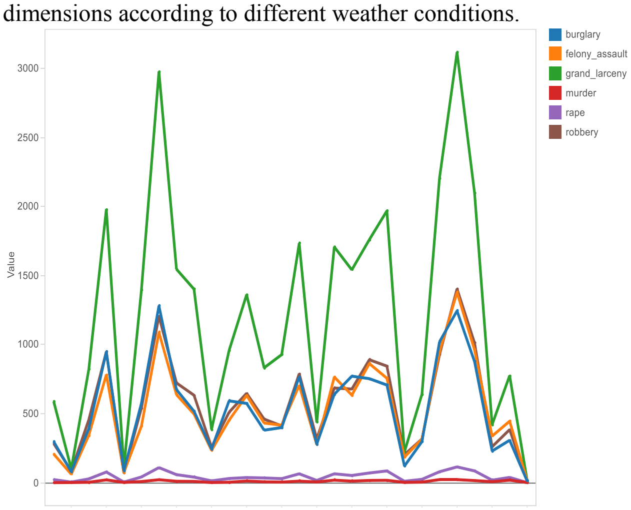

Fig. 1. Different kinds of crimes through mean temperature dimension

The number of crimes seems to has a peak at 76 ºF, and a trough at 38 ºF. The distributions of number of different crime types followed the same pattern through the temperature dimension. But to find out the effect of temperature on crime number, the distribution of temperature should also be identified:

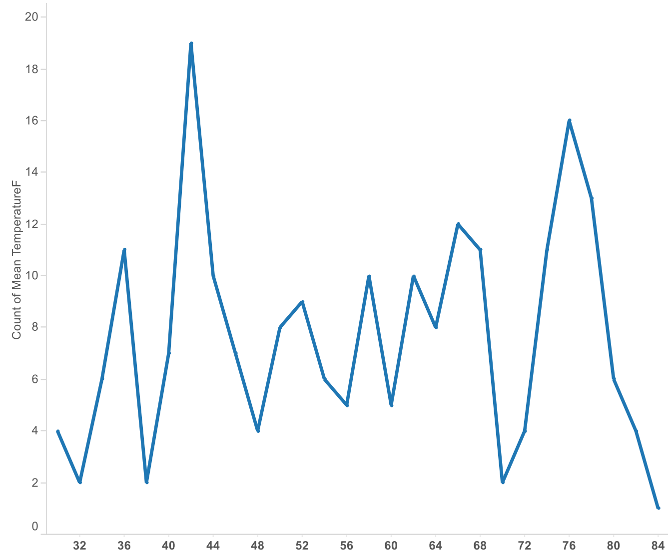

Fig. 2. Count of days in different temperatures in 2012

It's observed that the peak occurred at 42 ºF. Compared to Fig 1, it seems that although there're less days with higher temperature, but more crimes happened.

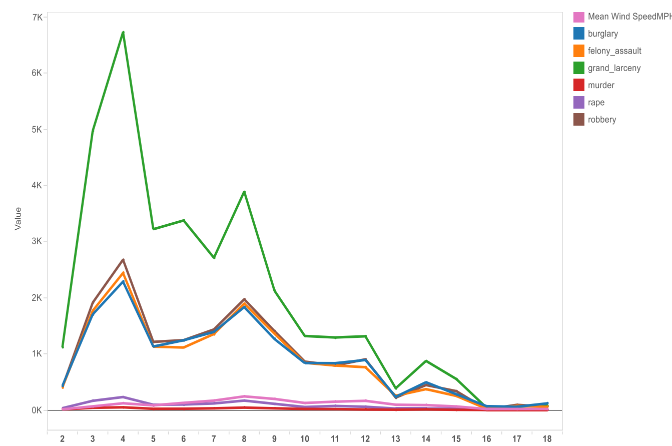

Fig. 3. Different kinds of crimes through mean wind conditions

The number of crimes has a peak at 4, and a trough at 10 MPH. The distributions of number of different crime types followed the same pattern through the wind dimension. The distribution of wind is also identified:

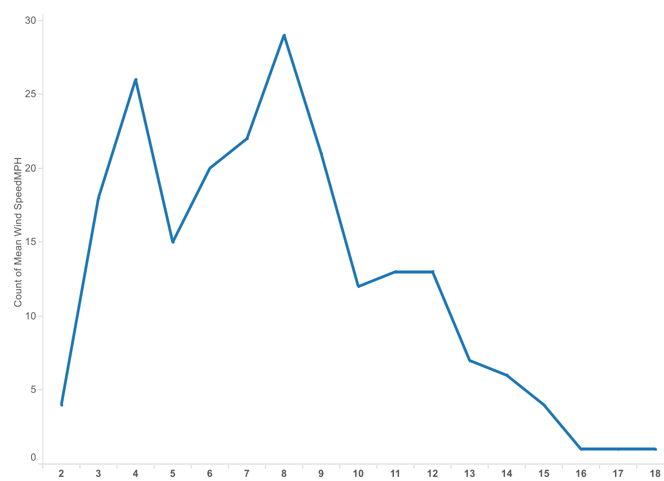

Fig. 4. . Count of days in different wind conditions in 2012

The count of days in different wind conditions has a peak at 8 MPH, compared to Fig 3, days with lower wind speed seems to have more crimes.

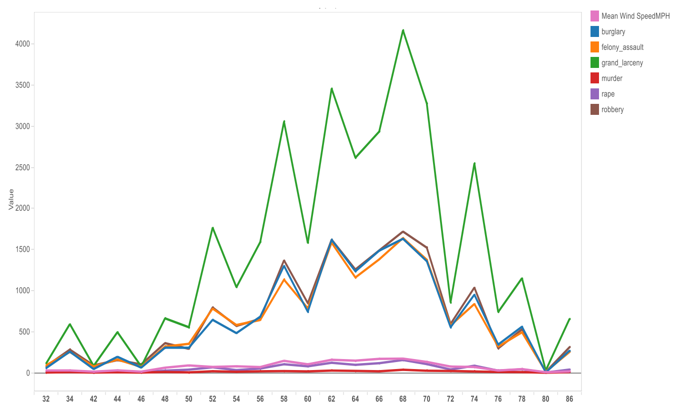

Fig. 5. Different kinds of crimes through mean Humidity conditions

The number of has a peak at 68 percentage, the distributions of number of different crime types followed the same pattern through the humidity dimension. The distribution of humidity is also identified:

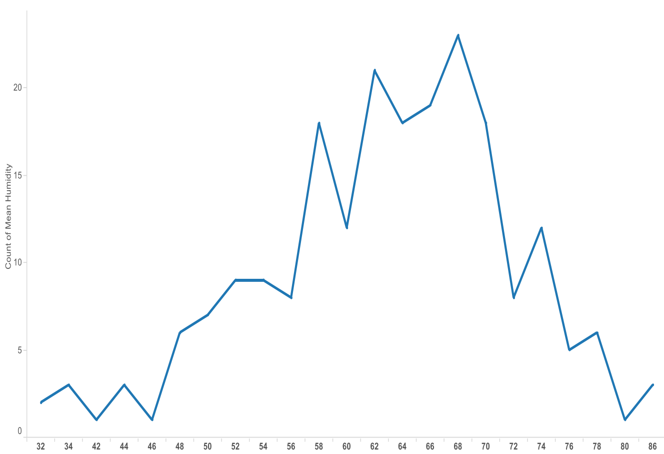

Fig. 6. Count of days in different humidity conditions in 2012

The count of days in different humidity conditions seems fellow the same pattern as the number of crimes. There's might no obviously effect form the humidity on crimes.

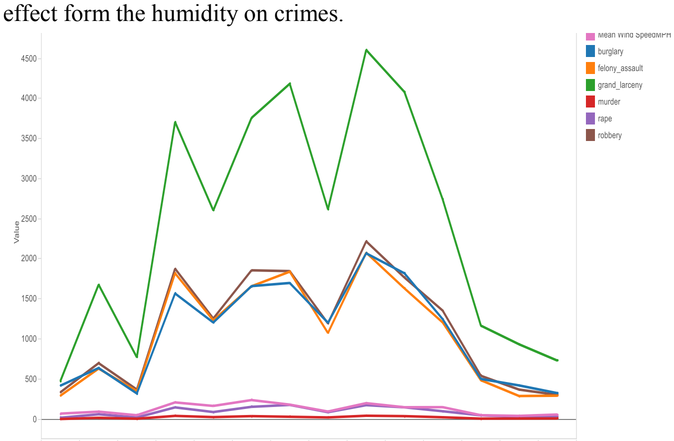

Fig. 7. Different kinds of crimes through mean sea press conditions

The number of has a peak at 30.05 inch, the distributions of number of different crime types followed the same pattern through the humidity dimension. The distribution of sea level pressure is also identified:

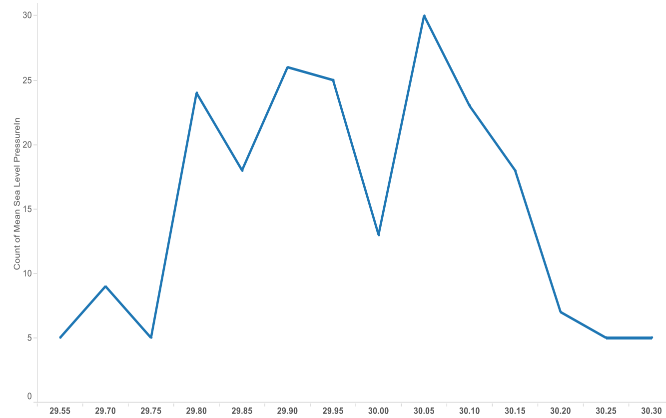

Fig. 8. Count of days in different sea level pressures in 2012

The count of days in different sea level pressures fellow exactly the same pattern as the number of crimes. There's might no obviously effect form the sea level pressures on crimes.

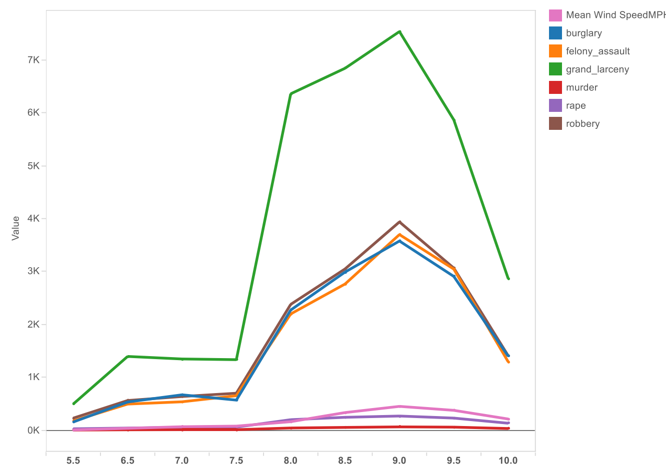

Fig. 9. Different kinds of crimes through mean visibility conditions

The number of has a peak at 9 miles, the distributions of number of different crime types followed the same pattern through the humidity dimension. The distribution of sea level pressure is also identified:

Fig. 10. Count of days in different visibility in 2012

The count of days in visibility conditions fellow exactly the same pattern as the number of crimes. There's might no obviously effect form the visibility on crimes.

The temporal patterns of the crimes are also explored:

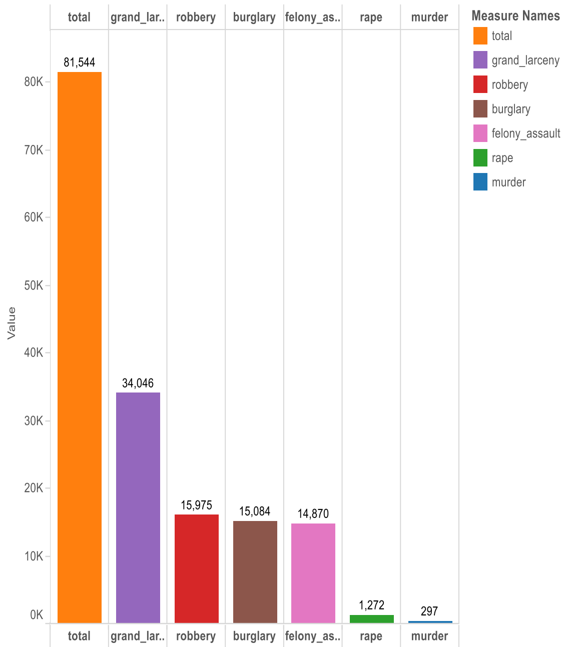

Fig. 11. Total number of each type of crime in city scale

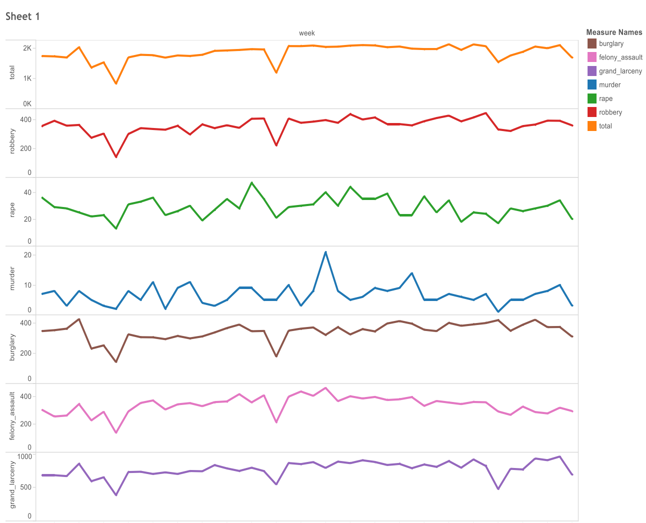

Fig. 12. Temporal pattern in city scale (separate, total number)

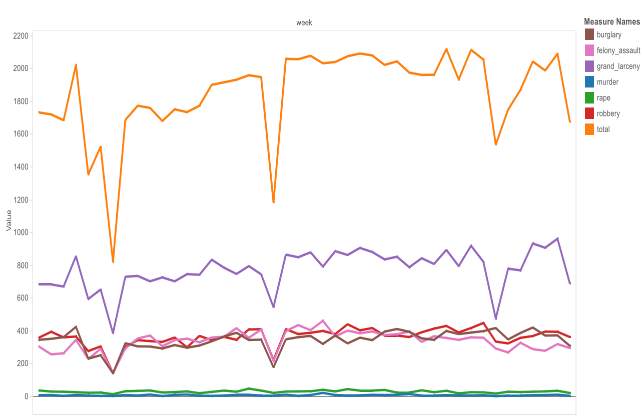

Fig. 13. Temporal pattern in city scale (combined, total number)

It’s obviously that grand larceny has the largest crime number in the city, robbery, burglary and assault are all around 15,000, rape is around 1300, and murder is the lowest one. Grand larceny, robbery, burglary and assault are following the same pattern, have obviously drop in the week 9, week 23 and week 44. In the former two drops, each of Grand larceny, robbery, burglary and assault has a drop, but in the last one, burglary has a slightly up. The last drop is different from the former two.

After checking the database, it’s found that week 9 and week 23 have no crime records of Bronx and Staten Island, this might be able to explain the first two drops, but the last one have the records of all the boroughs. Averaged crime number and weather conditions in weeks, then the last drop is found corresponding to a wind data peak, actually week 44 is the time hurricane Sandy hit. A possible conclusion is that the hurricane Sandy caused the drop the crime number in the week 44.

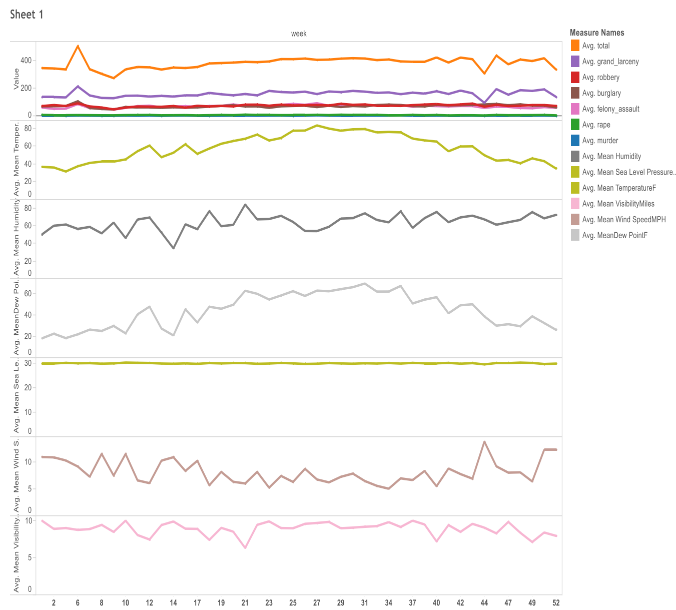

Fig. 14. Averaged crime number and weather conditions in weeks

The week 6 has the largest averaged number of crimes, but no obviously explanations can be found through the figures.

Linear regression model is used to find the quantitative relationship between different type of crimes and the weather conditions. Only robbery can be built an equation through the weather conditions.

Table 1. Robbery and weather conditions linear regression

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	（Constant）	-1594.3	907.64		-1.757	0.083
	Mean wind speed	13.932	4.123	0.897	3.380	0.000
	Max wind speed	-5.282	2.476	-0.547	-2.133	0.002
	Min Temperature	8.064	1.155	3.004	6.980	0.000
	Max Temperature	-6.617	1.103	-2.622	-6.002	0.000
	Max Humidity	-1.588	0.467	-0.381	-3.402	0.001
	Mean Sea level press	58.82	28.9	0.233	2.035	0.45

According to the linear regression of robbery and weather conditions using stepwise method, mean wind speed, min temperature, and mean sea level press seems to have positive relationship with robbery, the constant, max wind speed, max temperature and max humidity have negative relationship with robbery. These are hard to observe from the averaged crime number and weather figure. After checking the linear relationship between robbery and every weather condition, it's found that the linear charts are kind of misleading in Tableau and there are complete different feelings when the x or y axis's scales are changed.

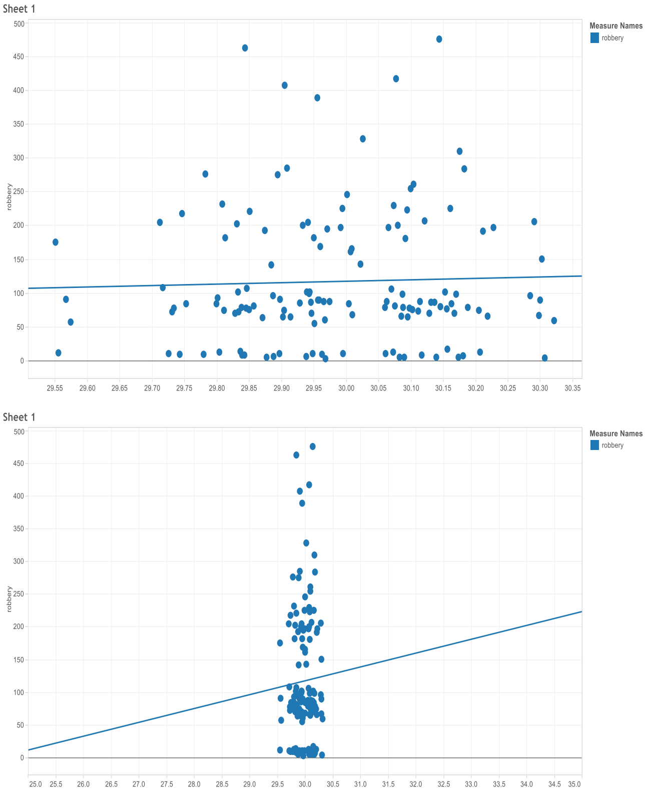

Fig. 15. Compare of different x axis scales through robbery and Mean sea level pressure

The linear regression model's R^2 for robbery is 0.55, the weather conditions has a high explanation ability to robbery. The model for robbery in the city scale is $Y = -1594.3 + 13.932*X1 - 5.282*X2 + 8.064*X3 - 6.617*X4 - 1.588*X5 + 58.82*X6$, where Y, X1, X2, X3, X4, X5 and X6 stand for number of robbery, mean wind speed, max wind speed, min temperature, max temperature, mas humidity and mean sea level press.

3.2 Borough scale

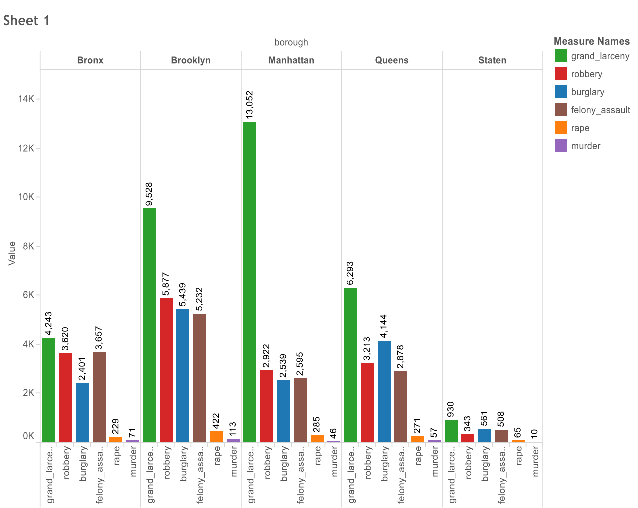

Fig. 16. Spatial crime pattern with different boroughs

Through the figure16, it's very easy to find out which borough has the highest number of each crime, like Manhattan has the highest grand larceny, Brooklyn has the highest robbery, rape, murder, assault and burglary.

However, both stepwise and enter methods can't build the regression equation at borough scale.

3.3 Precinct scale

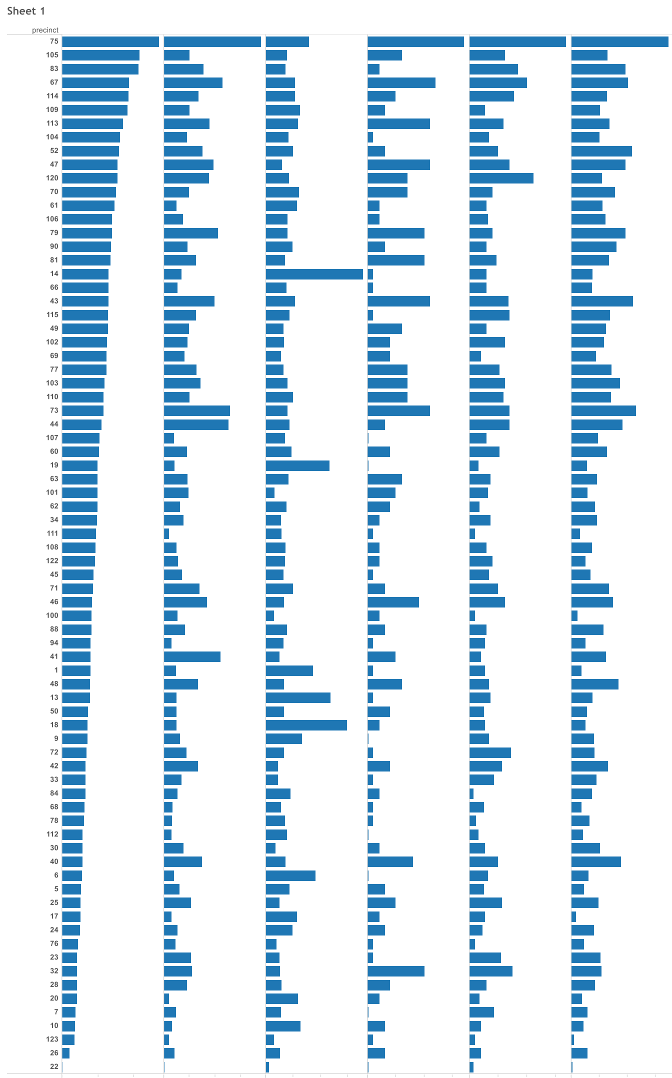

Fig. 17. Spatial crime pattern with different precincts

In figure 17, each row represent one precinct, each column represents one type of crime, from left to right are burglary, assault, grand larceny, murder, rape and robbery. The precincts are sorted according to their number of burglary from largest to smallest. Precinct 75 has the highest burglary, assault, murder, rape and robbery, and precinct 14 has the highest grand larceny.

The same as borough level, both stepwise and enter methods can't build the regression equation at precinct scale.

4 Conclusion

Through the visual analyses of crime and weather data in New York City, a number of significant observations are obtained: there seems no simple linear relationship between temperature and robbery; mild wind days have more crimes; an influence from the hurricane Sandy could be seem from the city scale analysis. The predication equation can be built at city level, but can't be built at borough or precinct level. And the most dangerous weather in New York City is temperature 42 ºF, wind 4MPH, humidity 68, mean sea level pressure 30.05 and visibility 9 miles in 2012; the crime facts of boroughs shows that Brooklyn is still the most dangerous place in the New York, the most dangerous precinct is precinct 75 according to the 2012 crime and weather data.

However, there are still several problems should be noted in this study: firstly, the data used in this paper is not perfect, there's no daily crime information available, and some weeks' data are missing which might cause a big difference in the findings, more samples are needed to draw any solid conclusion; secondly, predication model is simple, these can be greatly improved by various statistical methods, like adopt k-fold cross-validations to investigate models with higher dimension; thirdly, the usage of visual tools in this paper is still limited, Tableau can develop more attracting and useful visualizations with further explorations, not to mention there are numerous visual tools could be used in data analysis.

To implement a more comprehensive study, the first thing needs to be done is collecting more detailed and reliable crime and weather data. Once the daily and location information included crime data achieved, other more exciting approaches like GIS-based visual analysis would be obtainable. Another way to improve this study is to combine the temporal and spatial distribution information together, like looking into the change of crime number in different boroughs or precincts when the temperature goes from 30ºF to 40ºF. With all of these and other potential improvements, a more complete study is accessible in the future.

Acknowledgments

The author would like to thank Professor Enrico Bertini for the instructions and encouragement throughout the course of this work. The author also wish to thank all the colleagues in the Visual Analytics Fall 2013 class for their valuable comments and discussions.

References

[1] "Edwin Grant Dexter". Available at http://en.wikipedia.org/wiki/Edwin_Grant_Dexter.

[2] Edwin Grant Dexter. Weather Influence. London: The Macmillan Company, 1904.

[3] Ellen G. Cohn. Weather and crime. British Journal of Criminology, 30, 1990.

[4] Keith D. Harries, Stephen J. Stadler. Heat and violence: new findings from Dallas field data, 1980-1981. Journal of Applied Social Psychology, 18, 1988.

[5] Paul Butke, Scott C. Sheridan. An analysis of the relationship between weather and aggressive crime in Cleveland, Ohio. Weather, Climate & Society, 2, 2010.

[6] PediacitiesNYC. Available at http://nyc.pediacities.com/.

[7] Weather Underground. Available at http://www.wunderground.com/.