5-2016

GIS and Location-Based Crime Risk Analysis: Summer Internship with Location, Inc.

Zhilan Deng
zdeng@clarku.edu

GIS AND LOCATION-BASED CRIME RISK ANALYSIS:

SUMMER INTERNSHIP WITH LOCATION, INC.

ZHILAN DENG

DEGREE WILL BE CONFERRED MAY 2016

A GISDE FINAL PROJECT PAPER

Submitted to the faculty of Clark University, Worcester, Massachusetts, in partial fulfillment of the requirements for the degree of Master of Science in Geographic Information Systems (GIS) for Development and Environment in the department of International Development, Community and Environment.

And accepted on the recommendation of
Jie Tian, Academic Advisor

ABSTRACT

GIS AND LOCATION-BASED CRIME RIST ANALYSIS: SUMMER INTERNSHIP WITH LOCATION, INC.

ZHILAN DENG

My internship with Location, Inc. took place from May 20th, 2015 to August 24th, 2015. I worked with one direct supervisor, Jonathan Glick, as well as the CEO of Location, Inc. Andrew Schiller. I have four main responsibilities during the summer: 1) collecting U.S. crime point data, 2) geocoding and processing crime point data; 3) collecting and processing Canada Crime statistics and demographic data; 4) updating school performance data and U. S. crime statistics. This report includes the introduction of Location, Inc., where I did my internship, the details of my responsibilities in Location, Inc., and my assessment to Location, Inc.

Jie Tian, Ph. D.

Academic Advisor

ACADEMIC HISTORY

Name:	**Date:**
Zhilan Deng	**December 2015**

Baccalaureate Degree:

School of Resource and Environmental Science

Source:	**Date:**
Wuhan University, People's Republic of China	**May, 2014**

ACKNOWLEDGEMENT

I would like to thank Dr. Yelena Ogneva-Himmelberger and Dr. Jie Tian for their kindly help and great support throughout my time in Clark University.

For Dr. Yelena Ogneva-Himmelberger, thank you for giving valuable opportunities, kindly support and helpful suggestions in my work and life. You are also being an outstanding example for me by being always organized and critical as well as indulged in your projects.

For Dr. Jie Tian, thank you for giving me as well as all the international students in GISDE program warmly guide and support during our time in Clark University. Thank you for giving me suggestion both on academic work and personal life.

Finally, I would like to thank Jonathan Glick and Andrew Schiller, who gave me the the good opportunity to work with you in this summer. Thank you for giving me advice and making this internship both challenging and rewarding.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 1. INTRODUCTION

During my four years of undergraduate study in GIS in Wuhan University and one year of graduate study in GISDE in Clark University, my interest lies mostly in social application of GIS methods instead of environmental application of it. The Research Assistantship of Byrne project with Professor Yelena Ogneva-Himmelberger, which is a still ongoing project on gang activities and crime incidents in Main South, Worcester, from January 2015 to October, 2015 emphasized my interest in using GIS analysis methods and spatial data analysis in social application to improve residents' daily lives. Under this situation, I found the internship posting information of Location, Inc., a company which uses spatial data to create products such as crime risk report and neighborhood evaluation.

Job information is from Professor Robert Pontius' gisjob list (gisjobs@lists.clarku.edu), which sends information of GIS jobs to all the students enrolled in the list. After I applied to the position on their website, I was contacted by their leading geographic statistician, Jonathan Glick, the next day and scheduled a skype interview with him. About a week after the first skype interview, we scheduled a second interview in their headquarters at 86 Shrewsbury Street, Worcester and received an example work before I finally got this internship.

The internship in Location, Inc. is my first priority among other internships I have applied. It is a good fit of my interest of using spatial statistics, especially crime data, and my overall career goals of applying GIS knowledge into social application. Besides, the people in Location, Inc. are not only people with GIS background, but also business people and operation people who have limited knowledge of GIS, which will give me an opportunity on working and communicating with people in other areas.

My responsibilities in Location, Inc. in this summer is mostly searching public crime point data and processing them, as well as helping them updating the data they already had, which will be discussed in details in Chapter 3. I will also discuss about my assessment of this internship in Chapter 4.

CHAPTER 2. DESCRIPTION OF ORGANIZATION

2.1 Organization Introduction

Location, Inc. is a middle-sized SaaS (Software as a service) company that focuses on building and delivering big data about locations. The emergence of the company's main idea of this company is from a napkin in a restaurant in Portland, Maine, where Andrew Schiller put down his thoughts of location-based services in the year of 2000. From then, there are more than 40 million people utilizing Location, Inc. SaaS Products for their major investment, mitigating risk, protecting company budget resources.

The SaaS products of Location, Inc. include Crime Risk Assessment, Fraud Risk Assessment, Vehicle Collision Risk Index, and Real Estate, House Values, & Appreciation Rates, School Data Ratings, Lifestyle and Demographic Data, Neighborhood Reports for any given address in America, and Site Selection Search Engine. The goal of this company is to: build the best location-based technologies possible to empower businesses to make informed decisions that translate to increased revenue and reduced costs.

2.2 Location

Location, Inc. headquarters is located in Worcester, Massachusetts, the 2nd largest city in New England (Figure 1). The headquarters, located in the middle of Worcester's "Restaurant Row" which is famous for the great number of highly evaluated restaurants and bars, is yards away from Union Station and highway access. Most of the employees work at the office located in 86 Shrewsbury Street, Worcester, and some of the employees work remotely from Seattle and Canada.

2.3 Area of Expertise

There are three main areas of the expertise of Location, Inc.:

1) School Ratings: provides the quality of education in the given area based on government agency data.
2) Crime Rates: provides the crime risk at neighborhood level and discovers the safest neighborhood in any given city based on FBI crime reports.
3) Appreciation Rates: provides the information of the neighborhood with the highest home appreciation rates in a given area.

Currently, Location, Inc. focuses on providing services in U.S. and has expanded their services to Canada.

The customers of Location, Inc. include both companies and individuals who are interested in using the Location, Inc. SaaS products to help them make essential decisions.

2.4 Company Culture

The work in this company is both challenging and rewarding. This company keeps a good balance of life and work. The employees work hard to serve their customers' needs but can also relax by having a rest in the lounge in the office, having the beverages and snacks provided by the company, or taking a walk to the nearby park.

An important part of Location, Inc.'s culture is the strong combination among their employees. There are generally four teams in this company: inside sales representatives, geographical statisticians, web applications developers and operations managers. The

details of these four teams are described in 2.6. Each of these four teams has been assigned different kind of work and works relatively independently from each other. To unite them together, this company has staff meetings every two weeks for each team to report their accomplished tasks and future plans. By doing these, every team is able to get a good knowledge of the work other teams focus on and is encouraged to accomplish their tasks. Besides, this company takes all the employees out for lunch or drinks on every employee's birthday and all the important holidays. Location, Inc. also provides a retreat once a year to bring all the employees together and provides more opportunities for them to communicate with each other and building networks.

By doing all these efforts, Location, Inc. is able to build a friendly atmosphere in their work place and strong connection among their employees.

2.5 Company Benefits

There is a good reason to work for Location, Inc. In addition to the standard paid holidays, sick leave, and bereavement leave allowed in Location, Inc., this company provides coffee, tea, soda and snacks all the time as well as free lunches and drinks on the birthdays of employees. In terms of the traffic, Location, Inc. provides the parking lots for all the employees who need to drive to work. In terms of the working environment, Location, Inc. provides headsets, monitors, independent offices and all the equipment needed at work.

2.6 Employee Organization

Most of the employees in Location, Inc. are full-time employees, two of them are part-time co-ops, and several of them are support staff. There is a good balance of male and female in the headquarters of Location, Inc., with 5 males and 5 females. Generally, as mentioned in 2.4, there are four teams in this company.

1) Inside sales representatives: focus on contacting companies or individuals who are interested in Location, Inc. SaaS products and providing sales and business solutions to them.
2) Geographic statisticians: focus on using location-based data and statistical analysis to provide solutions on real estate investigation, mitigate risk and major investigation, etc. This is where GIS and mapping is usually done. Geographic statisticians need to collect public location-based data, using GIS tools and mapping methods, such as geocoding and spatial statistics, to produce SaaS products and maps which serve the customers' needs.
3) Web application developers: focus on designing and maintaining the website, including mobile website of Location, Inc. as well as its aligned website: Neighborhood Scout.

2.7 Location, Inc. Strengths and Weaknesses

There are obvious strengths and weaknesses in Location, Inc.'s career. One of their strongest strengths is their unique products based on the geographic concepts. Their concept of location-based data at neighborhood level and map-based products outstand

their services among other similar companies, and enable their products to provide more convincible and informative suggestions. However, the ways of selling the products might be a current considerable weakness in their company. The main method of selling their products is through telephone calls and the low frequency of phone call success might influence their profit.

CHAPTER 3. INTERNSHIP RESPONSIBILITY

3.1 Department Introduction

In this summer, I worked in the Data Statisticians team in Location, Inc. Data Statisticians team is responsible for collecting, managing and analyzing geographic and real estate spatial data used to build Location, Inc. SaaS products, including school evaluation, crime statistics, and demographics statistics. Besides, this team is also responsible for updating the existing data every year when new data is published. For example, for school performances data and crime statistics in the Federal Bureau of Investigation (FBI), Data Statisticians team updates them every year when new statewide exams are taken and new tables are published by FBI. Another responsibility of Data Statisticians team is to work with Sales Representatives Team to provide business solution for companies or individuals who are interested in using Location, Inc. SaaS products, and to work with Web Application Developer Team to provide Web services, such as online crime risk report for Web users.

This team currently has one leader statistical analyst and one co-op, as well as some external companies which are paid to work on data scratching and data entry.

3.2 Communication

Since my advisor worked remotely from Virginia in this summer, I received my tasks and discussed with my advisor mainly through emails or skype. To communicate with other teams, mainly the Inside Sales Representatives Team and the Web Application Developer Team, we either talked to them by email or in person since every team had at least one

representative at the headquarters of Location, Inc. To communicate with the whole company, we presented our completed tasks in the employee meeting every two months. In that meeting, all the four teams and the chairpersons in Location, Inc. got together and communicated with the whole company about future plans.

3.3 Responsibilities in the Internship

There were several responsibilities in my internship: 1) collecting U.S. crime point data, 2) geocoding and processing crime point data; 3) collecting and processing Canada Crime statistics and demographic data; 4) updating school performance data and U. S. crime statistics.

3.3.1 Collecting U.S. crime point data.

Location, Inc. can use crime point data to: 1) test the accuracy of crime risk assessment (figure 2) produced by Location, Inc.; 2) extract the timestamps of the point data and analyze the relationship between crime density and different time of a day.

I collected the crime point data either from the agencies' websites or from directly requesting. To be geocoded into crime point data, these datasets need to include latitude/longitude information or detailed addresses information. The example of the detailed information of the crime point data is shown in figure 3.

In total, I collected crime point data in 21 places in U.S., including counties, cities, towns, and villages.

3.3.2 Geocoding and mapping crime point data

Geocoding is an essential process to transfer statistical datasets to shapefiles which enable Location, Inc. to do further analysis.

To geocode the datasets with latitude/longitude information, I imported them in ArcGIS and using Display XY Data tool (figure 4) to map them.

To geocode the datasets with detailed addresses instead of latitude/longitude information, I first used Geocode Addresses tool in ArcGIS (figure 5), then I used Batch Geocode website (figure 6) to get the latitude and longitude information by inputting addresses. For those points that cannot be geocoded by the first two methods, I searched them on Google map and got the latitude/longitude information from Google Map.

The geocoded crime point data need to be clipped to places boundaries to avoid the useless data outside the area of crime risk reports. Different agencies report their data differently. For example, the crime type of "murder or nonelected manslaughter" is reported as "murder" or "homicide" in different agencies, some agencies report the crime type of "arson" but some agencies do not. Because of all these differences in the reports of different agencies, we only selected the crimes within the 7 main crime types defined as "Violent Crime" and "Property Crime" in FBI's Uniform Crime Reporting (UCR) Program, which are murder or nonelected manslaughter, forcible rape, aggravated assault, robbery, larceny-theft, burglary and motor vehicle theft. We also renamed the crime types reported by the agencies

to these 7 main crime types to make them identical among different shapefiles. The example of the processed crime point data is shown in figure 7.

3.3.3 Collecting and processing Canada crime and demographic data

To expand their business area to Canada and to attract more clients in Canada, Location, Inc. wants to collect Canada crime and demographic statistic data.

Both the crime and demographic statistics in Canada were collected from Statistics Canada (http://www.statcan.gc.ca/start-debut-eng.html). Similar to the crime point data, we only focused on the 7 crime types defined as "Violent Crime" and "Property Crime" by FBI. To connect the crime statistics with the previously downloaded census subdivision shapefile, I also need to specify the name and numeric id of each subdivision in the shapefile. The example of the processed crime statistics in Canada is shown in figure 8 and the example of demographic statistics is shown in figure 9.

3.3.4 Updating school performance data and U. S. crime statistics

Location, Inc. updates their school performance data and U.S. crime statistics every year to update their evaluation model and get the latest result of the evaluation of the neighborhoods in U.S.

School performance data were collected from the Department of Education of each state separately. Every state has its own statewide test that every student in the state is requested to take and the results are requested to be published on the government's website, the examples of the statewide tests are the Alabama State

Testing (ACT Aspire) test in Alabama and the Massachusetts Comprehensive Assessment System (MCAS) in Massachusetts.

U.S. crime statistics are collected from the website of FBI. Most of the agencies reports their crime statistics to FBI every year and FBI will collect them, generate them in tables and publish the tables on their website. The latest year of their crime statistics tables is 2014.

3.4 Conclusion

Location, Inc. is a SaaS company, their products need to be maintained, updated and improved. Most of my responsibilities during the summer were to help improve and update the model previously built as well as collecting new data which enables them to expand their services and make new products.

CHAPTER 4. INTERNSHIP ASSESSMENT

4.1 Learned Skills

This internship in Location, Inc., is very different from any project I have done during class. The greatest difference is that, in this internship, I worked with real data instead of designed data in class. For example, when geocoding crime point data, the addresses used to geocoding are not cleaned, they could sometimes include police codes or showing intersections like "Gates Street & Main Street". For those addresses, ArcGIS is unable to geocode them, so I need to clean them and find other ways to geocode them, even by manually searching them on Google Map and extract their coordinates. During class, geocoding process usually takes less than an hour to finish, however, in this internship, I spent almost a month to finish geocoding all the crime point data for the 21 places.

Besides, during class, I worked with peers with GIS background and also presented to audiences with basic GIS knowledge, however, in this internship, I worked with businessmen or developers and our work was presented to public and the majority of them know limited GIS knowledge and have their special needs. So we need to explain things without using GIS terminology and focus on customers' special needs, such as selecting a specific analysis level (neighborhood, census tract, city, county, etc.).

This internship gives me a good opportunity to explore Excel and SPSS, since the data need to be transferred between Geographic Statistician Team and Inside Business Representatives Team. Almost all the data files, including crime statistics and demographic statistics, have to be exported from SPSS output or Shapefile dbf table to Excel file and saved in csv format.

In the process of searching data, I also learned the background knowledge of the census data both in America and Canada as well as how to find the data we need. For example, most of the data are stored in government website, crime data are stored in the website of Police Department, and pawn shops data are stored in the website of Division of Finance and Corporate Securities (DFCS).

4.2 Skilled Learned at Clark University

In this internship, I used the knowledge of geocoding and Excel operations to do most of the tasks. I was told that the reason why I was selected was because of my experience on working with crime point data and census data. Also, a good knowledge of ArcGIS is also important in processing data and producing maps in this internship.

This internship makes me have a better understanding of the area that I am interested in, which is the social application of GIS methods. Besides, this internship intrigues my interests of working with spatial data. It shows me what data can do and how the products in Location, Inc. are built from a simple spreadsheet to maps, web applications and business solutions.

4.3 Conclusion

This internship gave me a lot of skills that I would not be able to obtain from school. Besides the technique skills, including geocoding, geo processing, utilizing and reading

Excel and SPSS, I also learned how to work and communicate with people with limited GIS knowledge and received the background knowledge of census, crime and business.

I would recommend this internship to other IDCE students for these reasons:

1) This internship is well-paid and located in Worcester. It is relatively hard to find a paid internship, especially in Worcester area because of their limited number. This will save a considerable expenses compared to internships outside Worcester area.
2) This internship is well-designed for students who are interested in statistics analysis, especially spatial data. All the products this company developed are based on spatial data and statistics. It would be very helpful to students' future career in the area of analysis spatial statistics by watching and participating in what they have done and how they realized their goals.
3) This internship is implemented in a friendly and professional team. Location, Inc. just won the best workplace among small sized companies in Worcester, where people will bring their dogs in the office and share food. Besides, every team not only finish their tasks in high-efficiency but also communicates with other teams efficiently and timely.

Generally, this internship is outstanding in many aspects and I will recommend this internship to students in Clark University.

CHAPTER 5. CONCLUSION

My experiences with Location, Inc. in this summer is overall challenging and rewarding, as they described in their job description.

It gave me a better understanding of the census concepts and crime information as well as the skills of communicating and working with non-GIS people, which added to my preparation for the future professional world. The working environment is friendly and efficient. I really enjoyed the comfortable office in Worcester and their annual retreat time in Ogunquit, which brought people together in a historical house rented by the company. I also very enjoyed working with the people in Location, Inc. especially my supervisor, Jonathan Glick, who directly worked with me. They are always willing to help and working hard together to reach the company's goals.

Because of the advantages of this position in many aspects, I recommend this internship to other students in Clark University who are interested in using spatial statistics or applying GIS methods in social applications.

FIGURES AND TABLES

Figure 1. The headquarters of Location, Inc. in Worcester.

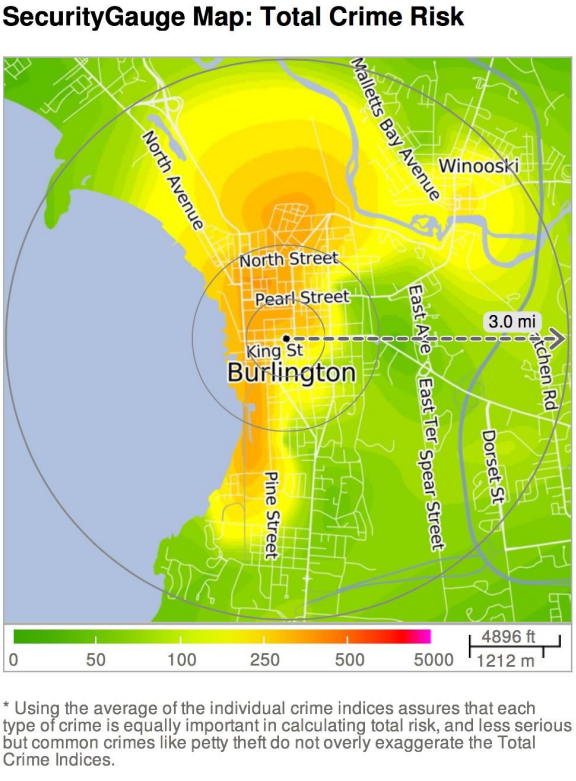

Figure 2. The Crime Risk Report.

Crime Data Sources	Filled by Zhilan					
URL for file download	**Has Lat/Lon coordinates for specific crime incedents**	**City Covered**	**Population**	**Crime Types Included**	**Time Period Covered**	**Other notes (optional)**
http://opendatacatalog.ashevillenc.gov/opendata/resource/35/city-of-asheville-crime-mapper-locations/	Yes	Asheville, NC	83,393	Aggravated Assault, Burglary, Drug Arrest, Larceny, Larceny of Motor Vehicle, Rape, Robbery, Vandalism, Homcide	1/1/2005 (complete) to 05/21/2015	
https://data.baltimorecity.gov/Public-Safety/BPD-Part-1-Victim-Based-Crime-Data/wsfq-mvij	Yes	Baltimore	620,961	Aggravated Assault, Arson, Assault By Threat, Auto Theft, Burglary, Common Assault, Homicide, Larceny, Larceny from	01/01/2010 00:50:00 to 02/21/2015 21:05:00	
https://data.cityofboston.gov/Public-Safety/Crime-Incident-Reports/7cdf-6fgx	Yes	Boston, MA	645,966	Aggravated Assault, GunShots, Aircraft, Argue, Auto Theft, Ballist, BeNoProp(?), BioThreat, Bomb, Burg Tools, Commercial	03/11/2012 6 am - 04/14/2015 8:38pm	Can be filtered and exported as csv, xls/xlsx/xml, and JSON.
Received from police agency	No	Boulder, CL	97,385	Theft, DUI, Shoplifter, Traffic, etc.	7/12/2013 to 05/31/2015	Data received in Excel format, includes X/Y coordinates.
Received from police agency	No	Bremerton, WA	37,729	ASSAULT - ARREST CHARGE', 'ASSAULT - ARREST CHARGE - DV', 'ASSAULT 1 - DEADLY WEAPON', 'ASSAULT 1 - DEADLY WEAPON - DV',	01/01/2003 to 04/22/2004	
Received from police agency	Yes	Burlington, VT	42,417	Airport AOA Violation, Alcohol Offense, Assault - Aggravated, Assault - Simple, Burglary, Domestic Assault - Misd.	8/1/2014 to 6/10/2015	Data received in Excel format, includes X/Y coordinates.

Figure 3. The example of the detailed information of crime point data.

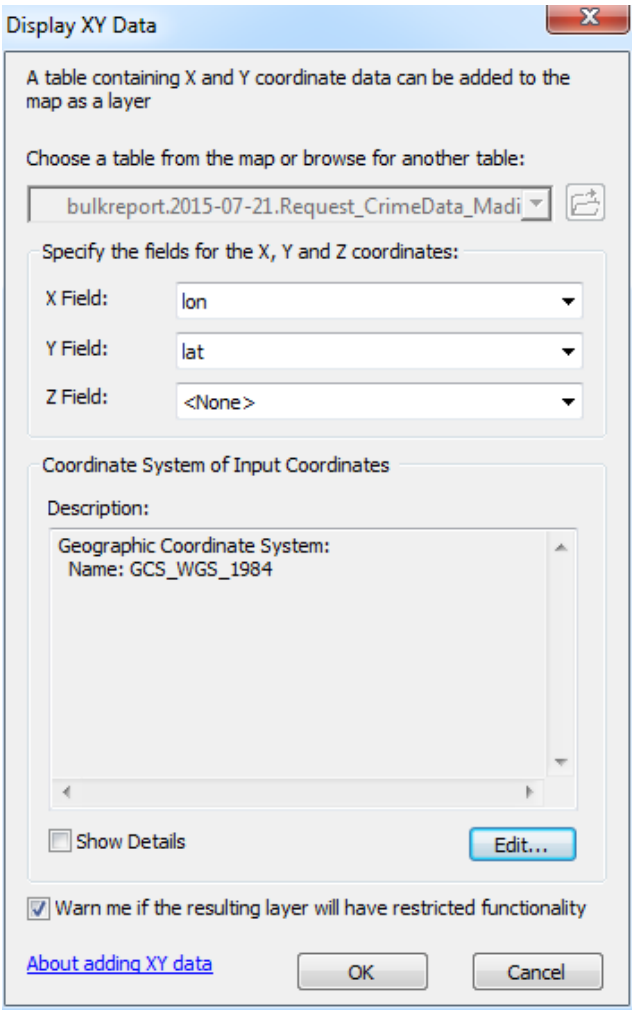

Figure 4. The interface of display XY Data tool.

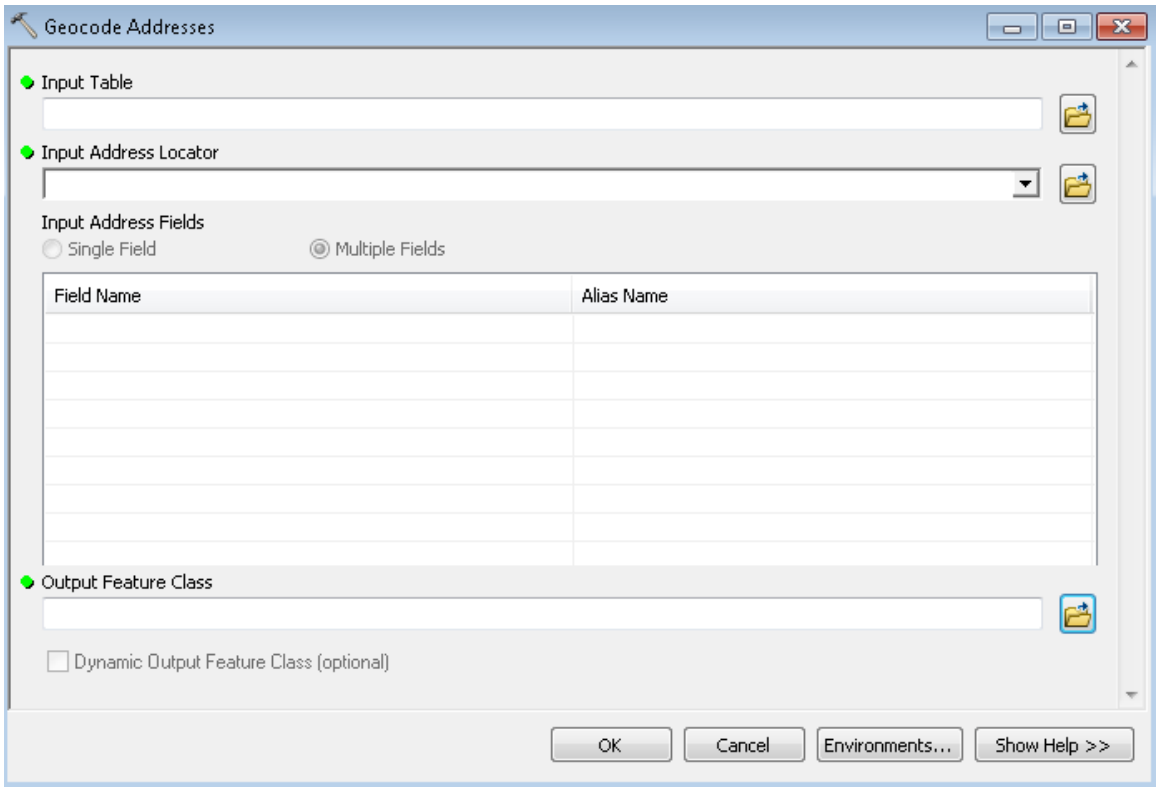

Figure 5. The interface of the Geocode Addresses tool in ArcGIS.

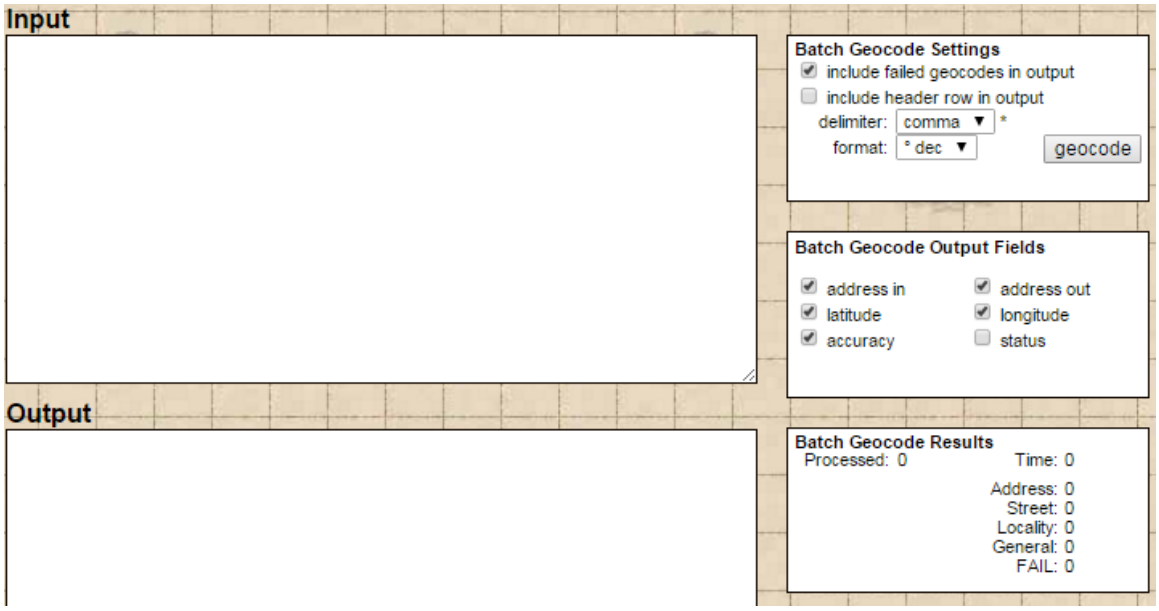

Figure 6. The interface of Batch Geocode.

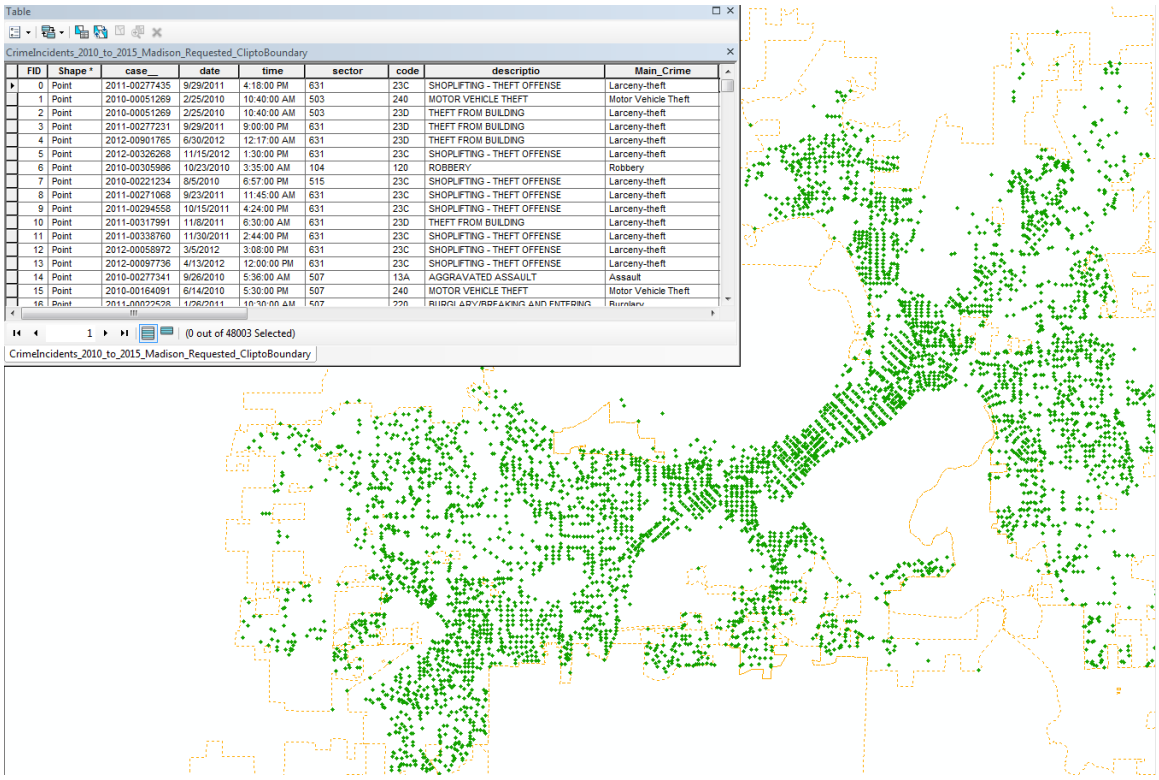

Figure 7. The processed crime point data in Madison, Wisconsin.

Municipal Police Area	Province	Homicide_2014	Sexual Assault_2014	Aggravated Assault_2014	Robbery_2014	Burglary_2014	Motor Vehicle Theft_2014	Theft_2014	Shapefile Name	Name-in-Shapefile ('CSDNAME')	ID-in-Shapefile
St. John's	Newfoundland and Labrador	1	124	280	131	1234	256	7558	Places_Matched MunicipalPoliceData_Newfoundland_and_Labrador	St. John's	1001519
Corner Brook	Newfoundland and Labrador	0	11	18	4	64	4	798	Places_Matched MunicipalPoliceData_Newfoundland_and_Labrador	Corner Brook	1005018
Labrador	Newfoundland and Labrador	0	11	4	0	24	3	78	Places_Matched MunicipalPoliceData_Newfoundland_and_Labrador	Labrador City	1010032
Montague	Prince Edward Island	0	1	3	0	5	3	100	Places_Matched MunicipalPoliceData_Prince_Edward_Island	Montague	1101018
Charlottetown	Prince Edward Island	1	19	32	17	139	18	2234	Places_Matched MunicipalPoliceData_Prince_Edward_Island	Charlottetown	1102075
Stratford	Prince Edward Island	0	3	6	0	27	1	210	Places_Matched MunicipalPoliceData_Prince_Edward_Island	Stratford	1102080
Cornwall	Prince Edward Island	0	0	1	1	20	2	164	Places_Matched MunicipalPoliceData_Prince_Edward_Island	Cornwall	1102085
Borden-Carleton	Prince Edward Island	0	0	0	0	0	0	0	Places_Matched MunicipalPoliceData_Prince_Edward_Island	Borden-Carleton	1103005
Kensington	Prince Edward Island	0	2	1	0	5	0	24	Places_Matched MunicipalPoliceData_Prince_Edward_Island	Kensington	1103014

Figure 8. The example of crime statistics in Canada.

Age group of child (13)	Total - Census family structure	Total couple families	Married couple families	Common-law couple families	Total lone-parent families	Female parent families	Male parent families
Total - Age group of child	43020	34520	32295	2225	8500	6895	1605
Under 15 years	25050	20640	19060	1580	4410	3660	755
0 to 4 years	8430	7200	6535	665	1240	1025	210
5 to 9 years	8155	6735	6255	480	1420	1190	235
10 to 14 years	8460	6710	6275	435	1750	1440	310
15 to 19 years	8815	6995	6600	400	1820	1450	370
15 to 17 years	5425	4250	3985	270	1175	950	230
18 years	1780	1430	1360	65	355	280	75
19 years	1610	1320	1255	65	290	225	65
20 to 24 years	5330	4375	4205	175	960	725	235
25 years and over	3820	2510	2435	70	1315	1070	245
25 to 29 years	1980	1525	1485	40	455	360	95
30 years and over	1840	980	950	35	860	710	155

Figure 9. The example of a demographic statistic table in Canada.