# A SPATIAL ANALYSIS OF CRIME IN THE CITY OF SAN LUIS OBISPO USING FREE AND OPEN SOURCE GIS SOFTWARE


By

Chad Bunn

Advised by Professor Greg Bohr


GEOG 461, 462

Senior Project


Social Sciences Department

College of Liberal Arts

CALIFORNIA POLYTECHNIC STATE UNIVERSITY


Winter 2014

# Table of Contents

# Senior Project Proposal: Crime Analysis of the City of San Luis Obispo

Chad Bunn

For my senior project I will be performing a crime analysis of the City of San Luis Obispo. I will be basing my analysis on reported crime data between January 1, 2012 and June 30, 2013 that were acquired from the San Luis Obispo Police Department. Crime analysis and mapping has become a necessity for law enforcement around the world and has proven to be a valuable resource in understanding criminal behavior and the environments that foster it.

To conduct the analysis I will be using varied open source GIS software including QGIS, CartoDB, TileMill, OpenStreetMap, and Leaflet. These open source tools will allow me to perform the analysis at no cost while also supplying me with tremendous amounts of support data. I will also be able to create beautiful and dynamic web based maps that the City of San Luis Obispo can use to show local crime patterns.

My analysis will include both spatial and temporal elements in order to establish accurate crime patterns. The City of San Luis Obispo experiences certain crimes more than others due to the local population of students and the homeless. Crimes associated with these two groups are expected in concentrated numbers in the downtown area and also in areas near the Cal Poly campus.

Annotated Bibliography
Chad Bunn

Works Cited

1. Ballatore, Andrea, Michela Bertolotto, and David C. WIlson. "Geographic Knowledge Extraction and Semantic Similarity in OpenStreetMap." *Knowledge and Information Systems* 37.1 (2013): 61-82. Print.

The authors of this article outline the benefits and processes in using crowdsourced geographic data. OpenStreetMap is an online map service that invites contribution by users with local knowledge of areas all over the globe. The article outlines the importance of establishing a defined naming system for all users to share to ensure that data is being added correctly.

2. Hart, Timothy C., and Paul A. Zandbergen. "Reference Data and Geocoding Quality Examining Completeness and Positional Accuracy of Street Geocoded Crime Incidents." *Policing* 36.2 (2013): 263-94. Print.

The authors of this article examine the use of geocoding, the process of locating positions based upon street addresses. There are many free programs that provide this tool, one of which CartoDB I will be using to geocode crime locations. The authors explain that an analysis of crime is only as good as the accuracy of the data being used.

3. Jackson, Shannon. "The Technology of Policing." *Technology and Culture* 51.2 (2010): 492-93. Print.

This article briefly explains the different technologies used in capturing crime data including CAD systems, GPS, and GIS.

4. Kounadi, Ourania, Thomas J. Lampoltshammer, Michael Leitner, and Thomas Heistracher. "Accuracy and Privacy Aspects in Free Online Reverse Geocoding Services." *Cartography and Geographic Information Science* 40.2 (2013): 140-54. Print.

This article examines the quality of geocoding services and the ramifications of open data and privacy. It asks the question whether or not there should be limits to open data.

5. Li, Weimin, and John D. Radke. "Geospatial Data Integration and Modeling for the Investigation of Urban Neighborhood Crime." *Annals of GIS* 18.3 (2012): 185-205. Print.

This article covers the topics of using GIS to determine crime patterns in an urban setting. The article discusses how an analysis of crime data can allow for future modeling of crime based on repeat victimization of certain locations.

6. Rice, Matthew T., Daniel R. Jacobsen, Douglas R. Caldwell, Scott D. McDermott, Fabiana I. Paez, Ahamad O. Aburizaiza, Kevin M. Curtin, Anthony Stefanidis, and Han Qin. "Crowdsourcing Techniques for Augmenting Traditional Accessibility Maps with

Transitory Obstacle Information." *Cartography and Geographic Information Science* 40.3 (2013): 210-20. Print.

This article discusses the growing presence of online maps that are open to the public as opposed to older methods that did not make data available in this way.  The authors argue that providing maps and data in this fashion is beneficial to the community it serves.

7.  Santos, Rachel Boba. *Crime Analysis and Crime Mapping*. Thousand Oaks, CA: Sage Publications, 2005. Print.

This was the first textbook written for the analysis and mapping of crime.  It details the challenges and processes that a new generation of crime analysts will use to determine crime patterns while using GIS.

8.  Steiniger, Stefan, and Andrew J.S. Hunter. "The 2012 Free and Open Source GIS Software Map - a Guide to Facilitate Research, Development, and Adoption." *Computers, Environment & Urban Systems* 39 (2013): 136-50. Print.

This article gives an overview of how to best use free and open source GIS data and tools.  I will be using several FOSS GIS programs and services to conduct my analysis of crime in SLO City.

9.  Vilalta, Carlos J. "How Exactly Does Place Matter in Crime Analysis? Place, Space, and Spatial Heterogeneity." *Journal of Criminal Justice Education* 24.3 (2013): 290-315. Print.

This article examines the spatial complexity of crime and how certain areas are more or less prone to crime due to their spatial qualities.

10.  Wang, Dawei, Wei Ding, Henry Lo, Tomasz Stepinski, Josue Salazar, and Melissa Morabito. "Crime Hotspot Mapping Using the Crime Related Factors--a Spatial Data Mining Approach." *Applied Intelligence* 39.4 (2013): 772-82. Print.

This article discusses the creation and use of hotspot crime maps to determine areas of high and low crime.  Hotspot maps are beneficial to determine where future crimes may happen and help law enforcement learn what is causing the spatial pattern.

Senior Project Outline:
Open Source Spatial Analysis of Crime in The City of San Luis Obispo
Chad Bunn
November 22, 2013

Introduction:

Geographic Information Systems have enabled law enforcement officials worldwide to study and analyze crime based on spatial properties. The City of San Luis Obispo is no different and employs GIS professionals to aid police officers in their work. The San Luis Obispo Police Department's freely accessible crime logs coupled with the growing Free and Open Source Software produced by the GIS community allows crime analysis to be done by local residents knowledgeable about these tools.

1. Spatial elements of crime that can be studied with GIS
    A. Location of reported crime as an address, a specific point with coordinates, a city block, a specific alley way, an area within a city, etc.
    B. Association of the crime location combined with environmental factors such as streetlights, bars, vacant lots, parking lots, blight, parks, schools, etc.
    C. Type of crime coupled with its location.

2) Temporal elements of crime that can be studied with GIS
    A. Time of day that crime was reported
    B. Day of week and month that a crime was reported
    C. Monthly crime reports, comparisons between months of previous years

3) GIS tools that allow analysis of crime data

A. Geocoding of addresses where crime was reported, creation of lat/long points from address information
B. Computer languages that permit easy cleaning of spatial data: Python, Ruby
C. Heatmap or Point Density tools that show hot spots of criminal activity through spatial algorithims
D. Geoprocessing tools that are able to associate points of crime with nearby features such as parking lots, schools, bars, etc. based on a radial distance.

4) FOSS GIS programs that incorporate these elements

A. Definition of FOSS GIS, Free and Open Source Geographic Information Systems
B. Benefits of FOSS GIS as compared with proprietary software including lack of cost, large community of helpful users, reliability and ease of use.
C. Software programs used for my analysis in sequential order: CartoDB for georeferencing and easy display of data for public, QGIS 2.0 for analysis of data generated from CartoDB, TileMill and Mapbox for creation of webmaps that the public can interact with.

5) Data acquisition and preparation for spatial analysis

A. Daily police logs from SLO PD published online in a .txt format with address, date, time of call, type of crime, time of dispatch, and time of clearance. Records Manager at the police department supplied me the police logs from January 2012 until June 2013 in a .txt format.

B. GIS software requires tabular data in order to perform necessary SQL functions and for general database design. The police logs were converted to .CSV or Comma Separated Values using a computer script written in the Ruby language. This script was written upon request after I entered an online form seeking help with data cleanup.

C. This data was then loaded into CartoDB, an online georeferencing tool in order to create point data representing each crime log record from the tabular data. From CartoDB the data was exported as a shapefile and ready to use in QGIS 2.0, the leading FOSS GIS software suite. Once in QGIS incorrectly georeferenced data was removed leaving me with a 90% rate of successful creation of point data.

6) Spatial analysis of over 45,000 records preformed in a variety of ways considering a variety of factors

A. Data split up by month with about 2500 records per month. This was done by applying SQL statements that would select the desired data only (“date” LIKE ‘2012-01-%’).

B. Generation of heatmaps for all 18 months using point density based on a 1000 ft radius around each point of crime resulting in a GeoTIFF image showing where point density through the value of individual pixels. From here I created polygons representing the hotspots for crime from each month using a built in tool that transforms raster data to vector data: Raster to Polygon.

C. Comparison between 18 months of data to establish consistent hot spots and outliers based on my data.

D. ~~Analysis of crime data based on proximity to environmental features such as parks, schools, bars, apartment complexes, etc. using buffer and intersect tools~~

E. Identification of repeat addresses over 18-month period including photos of and description of location.

7) Temporal analysis of over 45,000 records based on various fields

A. Day of week comparisons, weekday versus weekend comparison, monthly comparisons including summer monthly average crime rates versus rates of crime during months that Cal Poly is in session. Possible to do this with FOSS statistical software? GroupStats

B. ~~Possible association of spatial analyses and temporal analyses~~

**RESULTS—yet to complete**

A. Map of all data with city limits

B. Map of different heatmaps showing crime hotspots

C. Identification of high crime areas

D. ~~Identification of high crime days and hours~~

8) Creation of print map and user-friendly webmap to showcase the data and the findings of my analyses through the use of TileMill, Mapbox, and CartoDB.

A. Process of designing effective print map for display of crime data to public and city officials.
B. Process of creating beautiful and efficient webmap to dispay findings to public and FOSS GIS community.

Conclusion

-Success of analysis

-Success of webmap design

-Success of FOSS GIS tools to perform spatial operations

-Benefits of spatial analysis of crime in SLO City for police department and public.

# A Spatial Analysis of Crime in the City of San Luis Obispo Using Free and Open Source GIS Software

## Introduction

Advances in Geographic Information Systems (GIS) technology over the past decade combined with a growing community of users and developers have been of great importance to law enforcement agencies. Criminal behavior in an urban setting is readily studied in a geographic nature thanks to modern Computer Automated Dispatch (CAD) systems and GIS tools. These analyses are crucial to helping law enforcement understand current crime trends and identify hotspots of criminal activity. The findings from crime analyses are also important for the public to be aware of to ensure the highest level of safety possible to the citizens that a law enforcement agency is tasked with protecting and serving. Findings from crime analyses should be shared with the public in the simplest and most effective form in order to ensure the data is understood. Like many other law enforcement agencies, The City of San Luis Obispo Police Department logs and posts daily crime reports online for the public to view. However, due to the high cost of current GIS software and a limited budget the department is not able to present the data to the public in a user-friendly manner. Free and Open Source Software for Geographic Information Systems (FOSSGIS) makes this possible by removing the hefty price tag common to proprietary GIS software. The City of San Luis Obispo is a prime example of a city that can benefit from a FOSSGIS approach to crime analysis due to its population and limited budget and personnel. My analysis will show that large amounts of crime data can be processed and studied then later visualized and quickly shared using FOSSGIS and locally available data.

**Background**

Reports of criminal activity inherently possess many spatial characteristics that can be studied and located on the ground. When a citizen reports a crime they are required to tell the dispatcher the address or cross streets along with the type of crime that is being reported. This address represents a point on the ground which is associated with a pair of coordinates, latitude and longitude, which have previously been generated in an address point dataset. Once the address has been received the dispatcher runs it through the CAD system to alert officers and route them to the exact address of the crime being reported. If the citizen is unable to identify the address or cross streets of the crime they wish to report they can identify landmarks for the dispatcher to route officers to. These landmarks may be city parks, vacant lots, creeks, or other naturally occurring physical features. These reports are recorded and kept so that over time crime trends can be studied and spatial patterns can be better understood which informs officers on proper measures to take in order to deter crime (Santos, 2005). As law enforcement agencies map where types of crime are occurring specific hotspots of crime begin to emerge as environments tend to affect criminal behavior. Environmental factors such as streetlights, vacant lots, parking lots, areas of blight, schools, bars, etc. can affect criminal activity. Spatial characteristics are key elements of crime reports and when coupled with temporal attributes criminal patterns become even more evident and clear. The CAD system adds temporal data to each crime report including the date and time that a report was received, the time that officers were dispatched, the time officers arrived, and the time officers cleared the area. Temporal data allows officers to understand when crime occurs and as data is accumulated obvious patterns can emerge. These patterns can include what days of the week experience the most and least crime reports, which months of the year are busiest for law enforcement officials, and more specifically when specific types of crime are happening during the day, week, month, and year. Both temporal and spatial data are compiled and logged and then later disseminated to the public via the internet. The City of San Luis Obispo Police Department posts daily a log of crimes that occurred on the previous day, however with no accompanying map to give the public reference. Maps for reference can be found at the local newspaper's website, The San Luis Obispo Tribune, but it is a slow platform and does not allow users to easily understand where what types of crime are happening. My analysis will show the enormous benefits of using FOSSGIS to process crime reports and then generate easily understood print and digital maps.

```
PK
03/05/14              San Luis Obispo Police Department                     521
07:37                           Summary Report                         ============
140304009 03/04/14 Received:07:23 Dispatched:07:25 Arrived:      Cleared:07:25
===============================================================================
Type: ABAND VEHICLE                                              Location:LZ2
As Observed:
      Abandoned Vehicle

Addr: 1065 CORTEZ, San Luis Obispo, CA          Clearance Code:72HR TAG FOR 112

Responsible Officer: VanHooser, M
Units: 42K2
 Des: incid#=140304009 Completed call clr:RTF oc:ABAN call=91
CALL COMMENTS: SIL CHRYSLER 300
===============================================================================
140304011 03/04/14 Received:07:51 Dispatched:07:51 Arrived:07:52 Cleared:07:53
===============================================================================
Type: Found Property                                             Location:LZ2
As Observed:
      Lost or Found Property

Addr: 1299 NIPOMO & PACIFIC, San Luis Obispo, C  Clearance Code:Report To Follow

Responsible Officer: Hunter, S
Units: 42K3
 Des: (MDC) Completed call incid#=140304011 call=111
CALL COMMENTS: CDL MORGAN USILTON
===============================================================================
140304012 03/04/14 Received:08:28 Dispatched:      Arrived:08:34 Cleared:08:48
===============================================================================
```

**Figure 1.** The San Luis Obispo Police Department posts a log daily on their website http://www.slocity.org/, but with no accompanying map to give the public any reference.

**GIS Tools**

The current state of GIS tools and libraries allow users to manipulate and visualize data in limitless ways to better understand the world around us. Spatial analyses of criminal activity make use of many of these tools and libraries to turn addresses of reported crimes into features and then for these points to be studied in relation to each other and their surroundings. GIS tools refer to the tools that have been developed to allow GIS users to run complex computer codes that turn text essentially into spatial data. An important tool necessary for spatial crime analyses is one that transforms an address into a point with both latitude and longitude coordinates, more commonly known as geocoding. This tool turns the text depicting an address, the White House for example '1600 Pennsylvania Avenue' into 38.8977° N, 77.0366° W. The process entails dividing up a line representing a street into equal sections so that a range of values can be given to the line segment (Hart & Zandenberg, 2013).

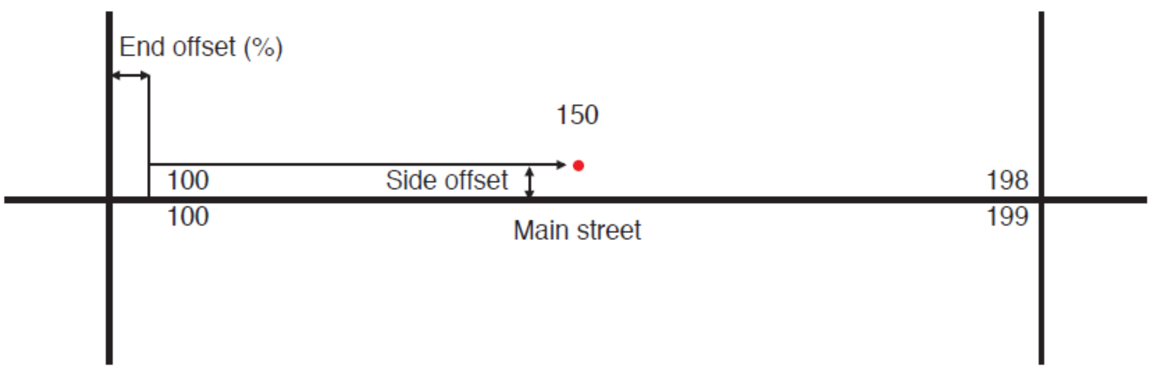


Figure 2. An example of the geocoding process (Hart et al., 2013) .

Another common tool used in identifying patterns of criminal activity is often known as point density. The point density tool analyzes the distance between all points in a dataset based on a designated distance and then locates the areas with the highest and lowest densities of points. The data behind the points allows us to understand where specific types of crimes are concentrated and what areas within a city are experiencing high crime rates. This tool is generally used to create heat maps to easily display the concentrations or hotspots of points. Another suite of tools known as geoprcessing tools are generally used to relate different features. Geoprocessing tools can identify locations where features interact, overlap, are within a certain distance of each other, and many other spatially based criteria. Geoprocessing tools are very important to spatial analyses because they reinforce Tobler's first law of geography which states that, "Everything is related to everything else, but near things are more related to each other" (Harvey, et al., 2012). GIS libraries are what houses the code for these tools and operate under a common programming language. A common language that GIS tools are programmed with is Python, which is an object oriented programming language. This means that Python is easily written to accommodate spatial data. Programming languages can quickly turn an unusable text file into spatially enabled data such as points, lines, or polygons. Together this suite of tools and libraries allows GIS users to perform complex analyses of anything with a spatial component including reported crime.

**FOSSGIS**

Geographic Information Systems have existed for decades and have evolved from strictly computer code into user friendly software packages that make analyzing and displaying data very easy and learnable. These GIS software packages have become invaluable to local, federal, and state governments and agencies because they allow another way for data to be analyzed and dispersed to the people whom they serve. Because demand is so high for these software packages, makers of popular GIS software can charge high rates to its customers. ESRI, a popular company in the GIS world thanks to its ArcMap package, charges users thousands of dollars for its most basic package. From there users are charged for additional tools and capabilities they wish to be included in their software. This creates an inefficient system where clients become wholly dependent upon ESRI for software to be improved or fixed while at the same time not being able to request many changes to the software. In the end it is not a democratic model of business, which is neither uncommon nor immoral but in our modern age where social media is king a better way has emerged. Free and Open Source Software for Geographic Information Systems is the idea that GIS software be both free of cost, though donations are required for it to thrive, and also changeable by users and developers. Open source software is software that has released its code to developers so that they make improve upon it and then have their improvements be incorporated back into the main software for all users to use. According to Steiniger and Hunter there are four primary freedoms associated with FOSSGIS : "(1) the freedom to run the software for any purpose (e.g., may it be education or business), (2) the freedom to study and adapt the software for own needs, (3) the freedom to redistribute the soft-ware, and (4) the freedom to improve the software and to release improvements to the public (Steiniger & Hunter, 2012)." Along with the appeal of a no cost software package and the ability to change the source code, there exists a large community of users dedicated to improving the software and helping each other learn how to more efficiently use it. The most popular FOSSGIS software package is QGIS, formerly Quantum GIS. QGIS has been in use for over a decade and within that time has experienced dozens of updates and improvements to get it to its current edition of QGIS 2.2 Valmiera. QGIS is similar to ESRI's ArcMap package in that it is able to run complex spatial analyses, visualize and style spatial data, and create printable maps. QGIS is easily downloadable from its website for users on all platforms including Mac and Linux, something ESRI is yet to offer, and users can contribute to and download a nightly version of the software that includes all the latest changes and improvements. Another benefit of QGIS is that users are encouraged to develop new tools that other users can install to improve upon the software. These tools are known as 'plugins' in QGIS and are similar to ArcMap's extensions which come at a cost to ESRI's customers. The existence of a FOSSGIS alternative is beneficial to the entire GIS community because it encourages competition with other for profit GIS companies causing them to improve their software and lower costs so that in the end the users are greater beneficiaries of GIS software. Barriers to firms adopting QGIS include the widespread use of ESRI specific file formats and a general lack of courses in GIS programs at colleges and universities that use QGIS to teach GIS. A detailed conclusion of my experience with QGIS in compiling this analysis will be included at the end of my report. Other elements of FOSSGIS include the availability of free data and also the availability of online platforms to publish shareable maps. A wealth of spatial data exists throughout the world, but access to it is not always open. OpenStreetMap aims to solve this problem by allowing users to create and extract data from anywhere in the world. OpenStreetMap is a volunteer collaborative online project to map the world with as much detail as possible. Anyone can join and begin digitizing the online map and anyone can edit your data in order to improve upon it. The data created by users is also then freely extractable by users to be used with GIS software (Ballatore, Bertolotto & Wilson, 2013) . As of March 5, 2014 there were 1,536,338 contributors to the project. Users are able to extract data representing many features for example: buildings, roads, trails, trees, land use designations, borders, lakes, rivers, parking lots, and almost any other physical feature that can be placed on a map. This is important in criminal analysis because a municipality's data can easily be shared with other interested users and users can contribute their local knowledge to provide a much more detailed and dynamic map. To further complement the FOSSGIS movement are web platforms designed to enable the sharing of wholly customizable and beautiful maps online. Companies such as CartoDB and MapBox offer free accounts that let users display maps they have designed. However as users require more storage space they must upgrade to paid accounts, though generally speaking for the average individual user the storage limits provided with the free accounts are sufficient. In order for me to perform my crime analysis I extracted data representing buildings, parks, parking lots, and schools in The City of San Luis Obispo from OpenStreetMap. I also acquired a .txt file from the San Luis Obispo Police Department of the daily crime logs from January 1, 2012 through June 30, 2013. To analyze the data and prepare printable maps I used QGIS 2.0 Dufour. And finally to display my maps and data I published several visualizations with CartoDB. The process was entirely free from start to finish and with the help of a larger community of FOSSGIS users it was very easy to solve any issues in my analysis.

**Data**

The City of San Luis Obispo Police Department not only posts a daily log of all crime reports recorded by their CAD system, but they keep a record of all reports ever generated by the system in a database. In order for me to preform my analysis it was necessary that I have access to a large portion of their dataset. The department's records manager, Kerri Rosenthal provided me with a text file composed of every crime report for the 18 months between January1, 2012 and June 30, 2013. Had I waited longer to do my analysis I could have received an even larger and more current dataset, but in the interest of time this selection of records was sufficient and current enough for my analysis. As shown in figure one previously the data came to me simply as a text or .txt file lacking any spatial reference like latitude and longitude coordinates. Data in this format is unusable in any GIS software package so it was necessary that each report be geocoded. CartoDB offers a free geocoding service that can be used to turn tabular data such as an Excel workbook into any of several types of spatial data including shapefile, GeoJSON, CSV, or KML. A tabular format allows for your data to be much more easily queried and styled with what is known as Structured Query Language or SQL. In order to convert my text file into tabular data I used Excel to separate all the fields into their own column and then populate them with their attributes. The text file provided me from the police department contained the following attributes: address, city, type, location, date, time received, time dispatched, time arrived, and time cleared. In order to separate or parse the text file I used the search and replace function in Excel to populate each field. This method was extremely time-consuming and I was only able to parse about a month's worth of data in an hour. To speed up the process I reached out on the popular online forum GIS.StackExchange.com. On November 13, 2013 I posted the following, "*I am trying to clean up a large .txt dataset in Excel. I think that a macro or another application may be able to clean up the data the way I want but I have limited programming skills. I need to create fields from different sections in the .txt file so I can then geocode the points. I am conducting a crime analysis for my local police department and up until now I have been using wildcards and the Find & Replace tool to do this, but I have nearly 2 years worth of police calls. Fields needed:"Address", "City", "State", "Type", "Location", "Date", "Day of Week", "Time Received", "Time Dispatched", "Time Arrived", "Time Cleared". I put up the data from January 2012 on Github in repo called SLO-Crime under my name, chadbunn. I will be using CartoDB to geocode the data and then QGIS to perform the analysis. I will probably make some maps in TileMill as well if I have enough time to. Any input would be appreciated because it is getting tedious! Thanks!*" Within less than 24 hours I had received several responses and possible solutions to my parsing problem. The solution I turned to came from a user only identified as 'toms' from Oakland, California. He had quickly written a script in the Ruby programming language that would separate my text file into the required fields and result in a tabular CSV file which I could then use with CartoDB. The script required me to load the Ruby programming language onto my computer and then I was able to run the script on my text file within a command line window. The script is listed below and it essentially separates out the fields I required and then populates them with their associated value from the text file. The script took less than two minutes to run through my year and a half of data and was immediately readable in Excel and on CartoDB.

```
 require 'date'
# open files
infile = File.new("CompleteData.txt","r")
outfile = File.new("result.csv","w")

# write header
header = "Address,City,State,Type,Location,Date,Day of Week,Time Received,Time Dispatched,Time Arrived,Time Cleared\n"
outfile << header

# read file lines
result = {}
infile.each do |line|
  values = line.split(" ")
  next if (values.length < 2 || line.chomp == "As Observed:")
  begin
    if values[0] == "Addr:"
      result["address"] = line.split(",")[0].gsub("Addr: ","")
    elsif values[0] == "Type:"
      result["location"] = values.pop.split(":").last
      values.shift
      result["type"] = values.join(" ")
    elsif values[2].split(":")[0] == "Received"
      result["date"] = values[1]
      result["day"] = Date.strptime(result['date'], '%m/%d/%y').strftime('%A')
      result["time_received"] = "#{values[2].split(":")[1]}:#{values[2].split(":")[2]}"
      result["time_dispatched"] = "#{values[3].split(":")[1]}:#{values[3].split(":")[2]}"
      result["time_arrived"] = "#{values[4].split(":")[1]}:#{values[4].split(":")[2]}"
      result["time_cleared"] = "#{values[5].split(":")[1]}:#{values[5].split(":")[2]}"
    elsif values[0] == "Des:"
      # write result
      outfile.write "#{result['address']},San Luis
Obispo,CA,#{result['type']},#{result['location']},#{result['date']},#{result['day']},#{result["time_received"]},#{result["time_dispatched"]},#{result["time_arrived"]},#{result["time_cleared"]}\n
"
      result = {}
    end
  rescue
    puts "Line was: #{line}"
  end
end

# close files
infile.close
outfile.close
```

**Figure 3.** The complete parsing script written by a user on GIS.StackExchange.com.

After having parsed my data and converted my .txt file into a usable tabular .csv file I uploaded it to my free CartoDB account. CartoDB offers a free geocoding tool that will create points based on the address in your dataset. CartoDB is able to do this because it makes use of Google's geocoding API, the same tool that will show you where an address is located if you search for a specific address on Google. This process took nearly an hour to complete, but it ran independently and in the end my tabular data was turned into 45, 801 individual points each with its own latitude and longitude. Google's geocoding API is sufficiently accurate for urban settings which meant it would work for my analysis of The City of San Luis Obispo. After geocoding my tabular data in CartoDB I was able to easily export it as a shapefile which is read by GIS software packages. Data exported from online mapping services such as CartoDB and OpenStreetMap use the Web Mercator coordinate system because it has a good representation of the entire globe. However, my data all falls within the more specific coordinate system of the North American Datum of 1983, in the California State Plane Coordinate System, Zone 5. In QGIS I was able to quickly assign my data to the local coordinate system and begin my analysis. As for the accuracy of my geocoded points, I found that my points were mostly concentrated within the city limits of The City of San Luis Obispo, with 90% of all the points within a three mile buffer zone of the city limits. This left me with 44,676 points in my dataset to be analyzed. This proved that CartoDB's geocoding service was accurate enough for my analysis as geocoding hit rates are generally rated as acceptable if they are above 85% accurate (Santos, 2005).

**Figure 4.** Each red dot represents a reported crime between 1/1/12 and 6/30/13. The Black dotted line represents the boundary of The City OF San Luis Obispo while the outer blue line represents a three mile buffer zone around the city's limits where 90% of the geocoded points were located.

**Analysis and Maps**

To conduct a spatial analysis of crime there are several factors to be considered including place, time, and type all of three of which were included in my dataset. To begin I organized my complete dataset by the top 50 occurring types of crime in order to accurately represent the types of crime that are most common to The City of San Luis Obispo. Originally there existed more than 100 types of reported crimes in my dataset, but for my purposes and in the interest of time I decided to only analyze the top 50 most common types of reports. This was quickly done by using the free QGIS plugin Group Stats that creates tables with basic mathematical functions, similar to those of Microsoft Excel, to organize data. After finding which types of crime were most common in my dataset I create individual shapefiles in order to produce density heatmaps to better display the data. My findings are shown in the table below:

| Number | Type | Count | Number | Type | Count |
|---|---|---|---|---|---|
| 1 | Suspicious | 3748 | 26 | Lost Property | 448 |
| 2 | Disorderly | 2924 | 27 | Fraud | 442 |
| 3 | Alcohol Offense | 2434 | 28 | Controlled Narc | 433 |
| 4 | Theft | 2177 | 29 | Public Works | 432 |
| 5 | Trespassing | 2095 | 30 | Assault | 426 |
| 6 | Alarm Audible | 2063 | 31 | Noise Other | 417 |
| 7 | Noise Party | 1815 | 32 | MC-Camping | 408 |
| 8 | Welfare Check | 1808 | 33 | MC Violation | 363 |
| 9 | Towed Vehicle | 1762 | 34 | Posting Vehicles | 338 |
| 10 | Assist Req | 1506 | 35 | Loitering | 336 |
| 11 | Abandoned Vehicle | 1421 | 36 | Juvenile Prob | 331 |
| 12 | Collision Non Injury | 1130 | 37 | Burglary Residential | 314 |
| 13 | Parking Problem | 1126 | 38 | Warrant | 307 |
| 14 | Traffic Offense | 1068 | 39 | MC-Sleeping | 306 |
| 15 | 911 Abandon | 1068 | 40 | Collision Injury | 296 |
| 16 | Traffic Hazard | 956 | 41 | Mental Subject | 287 |
| 17 | Found Property | 861 | 42 | Threatening | 237 |
| 18 | DUI | 860 | 43 | MC-Panhandling | 234 |
| 19 | Vandalism | 856 | 44 | Graffiti | 202 |
| 20 | Animal Problem | 603 | 45 | Attempt-Locate | 202 |
| 21 | Collision Hit And Run | 592 | 46 | MC-Alcohol | 200 |
| 22 | Noise Police | 590 | 47 | Burglary Vehicle | 195 |
| 23 | Assist of Police | 490 | 48 | Information | 189 |
| 24 | Citizen Dispute | 461 | 49 | Communications | 170 |
| 25 | Medical | 455 | 50 | Alarm Silent | 154 |

**Figure 5.** List of most common 50 crimes in my dataset in order of most common to least common.

Many of the fields are easily understandable, but those that begin with 'MC' signify a reported crime related to the city's municipal code. The top crime reported in my dataset is only vaguely identified as 'Suspicious' which refers to any activity that the person reporting it deemed to be as such. Those types of crime including the word 'Alarm' refer to security or fire alarms in buildings that when set off automatically alert the police department. From the list it is inferred that violent crime, defined in the FBI's Uniform Crime Reporting (UCR) Program as murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault, was not reported between January 2012 and June 2013. The most common groups of crime in the list those related to alcohol, traffic, the city's municipal code, alarms, and noise. After determining the most common types of occurring crime I generated density heatmaps for each which show where each type of crime was concentrated within the city. Below are some of the heatmaps I generated in order to display where the most common types of crime occurred between January 2012 and June 2013 in The City of San Luis Obispo. For reference I included a map of the city showing its streets, the main roads being labeled, and the city's parks, schools, and parking lots.

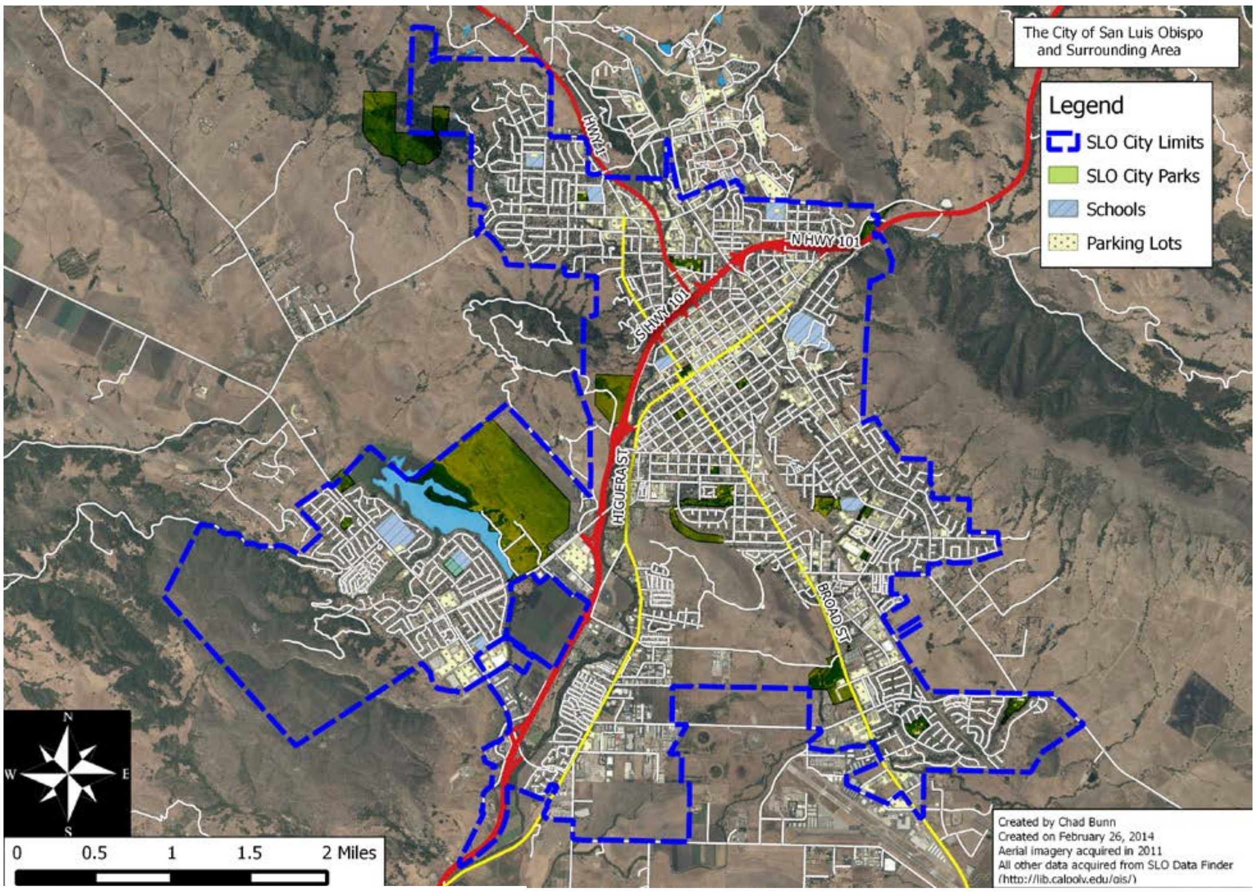


**Figure 6.** Map of The City of San Luis Obispo

The City of San Luis Obispo includes large swaths of open space, the downtown area at the heart, and both residential neighborhoods and commercial zones radiating outwards from there. Highway 101 divides the city into two halves.

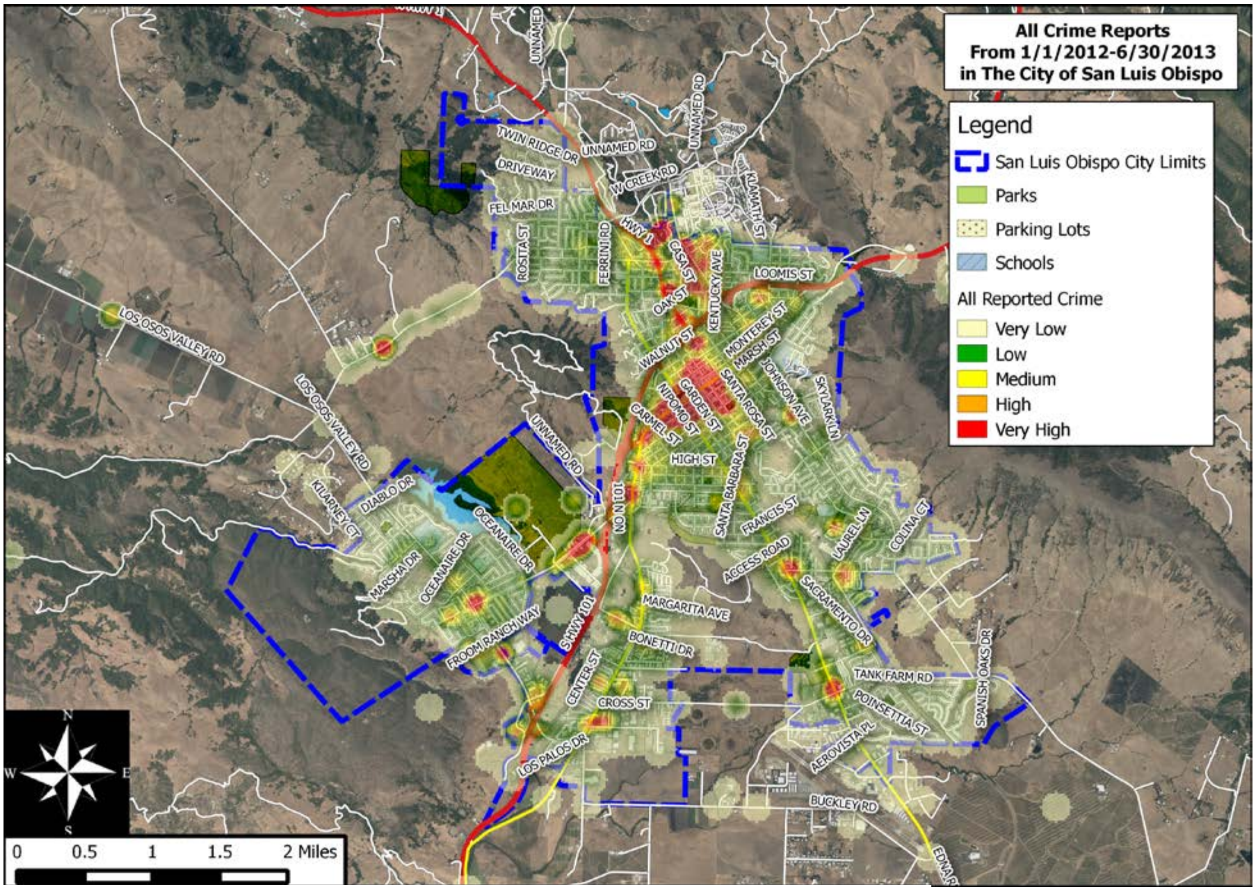


**Figure 7.** Map showing hotspots for all types in crime in The City of San Luis Obispo.

This map shows hotspots for all reported crimes from January of 2012 through June of 2013. The downtown area is clearly the largest hotspot for reported crimes with the north end of the city also showing a large hotspot. Smaller scattered hotspots appear throughout the city and tend to be located along the major roads of Broad Street, South Higuera Street, Madonna Road, and Los Osos Valley Road. This map merely serves to show where general hotspots occur but for more detailed maps I used the data I had organized by type of crime as this will show better patterns for individual crimes.

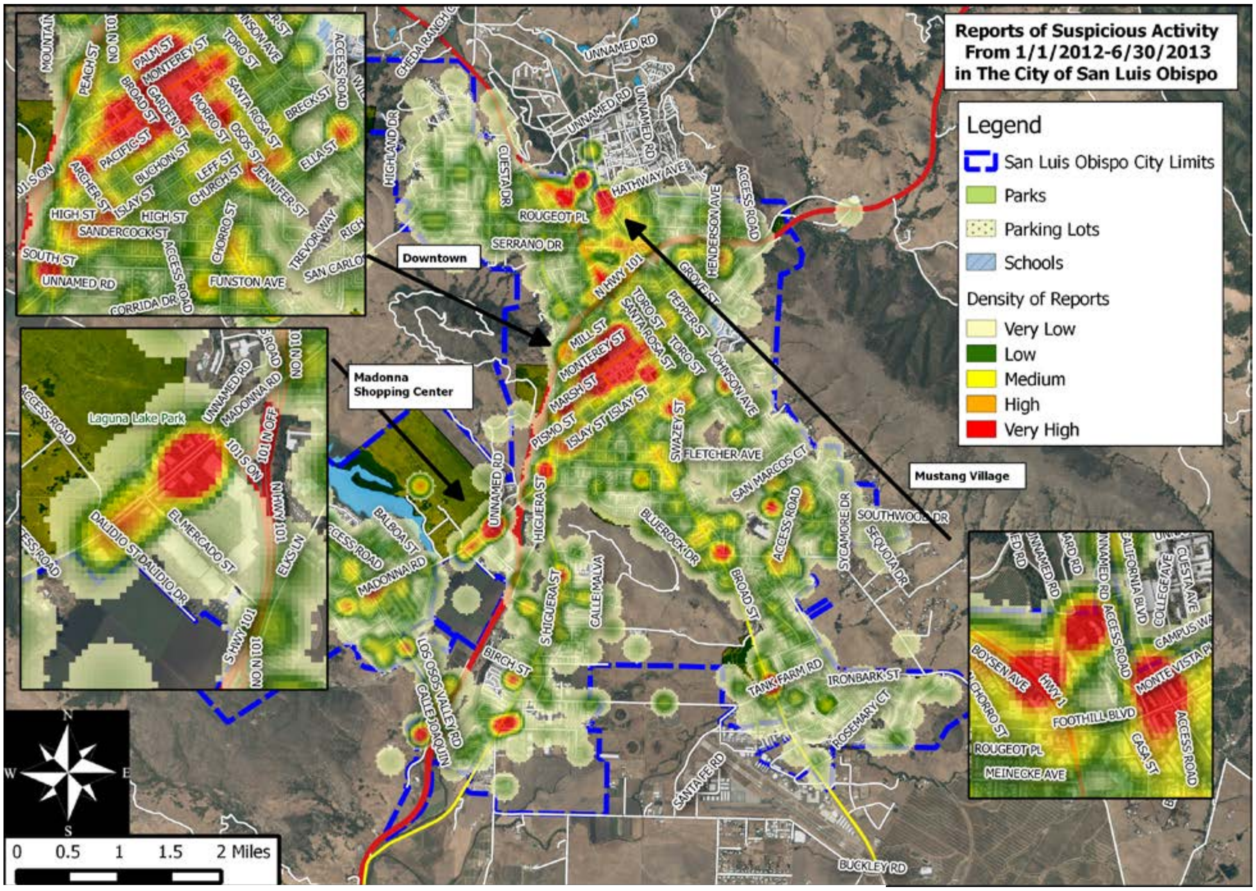


**Figure 7**. Map of all reports of suspicious activity in The City of San Luis Obispo.

Reports of suspicious activity were found to be concentrated in three main areas with some smaller clusters also appearing. The red zones on the map depict areas with a density of 0.39 reports of suspicious activity within a 500 foot radius. In the north of the city reports tended to be around the Mustang Village student housing near the campus of California State Polytechnic University of San Luis Obispo. Another hotspot appears in the downtown area of the city with Higuera Street running through the heart of the hotspot. One more hotspot appears across from Laguna Lake Park at the Madonna Plaza Shopping Center. These three hotspots contain are examples of three types of city zones: student housing, downtown commerce, and general commerce. General commerce in this analysis refers to commercial areas that are not specifically in the downtown area of the city. Reasons for these three hotspots for suspicious activity are clearly related to high concentrations of people. With high concentrations of people reports of all types are more common because there are more people witnessing crimes and then calling the police department. Analysis beyond this must be conducted at the ground level and cannot be readily achieved from solely a GIS standpoint.

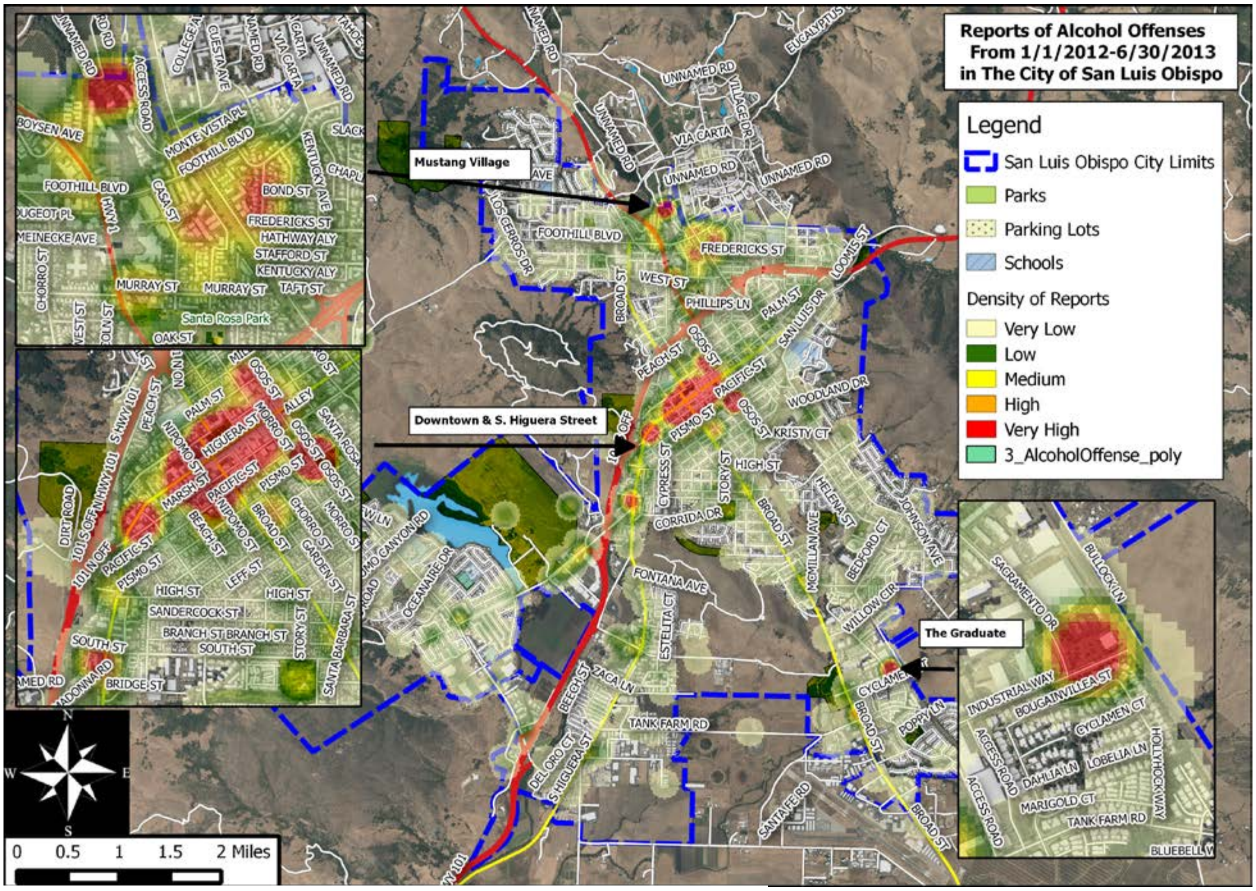


Figure 8. Map of all alcohol offenses in The City of San Luis Obispo.

Reports of alcohol offenses show a slightly different pattern than those of suspicious activity with three clear hotspots emerging: student housing at the north area of the city, downtown, and at nightclub The Graduate at the southeastern area of the city. Both downtown and The Graduate are examples of restaurants and bars attributing to the high density of alcohol offenses, while the area of student housing in the north is clearly the reason for high densities of alcohol offenses there. In this subset of data the red zones on the map depict areas with a density of 0.42 reports of alcohol offenses activity within a 500 foot radius.

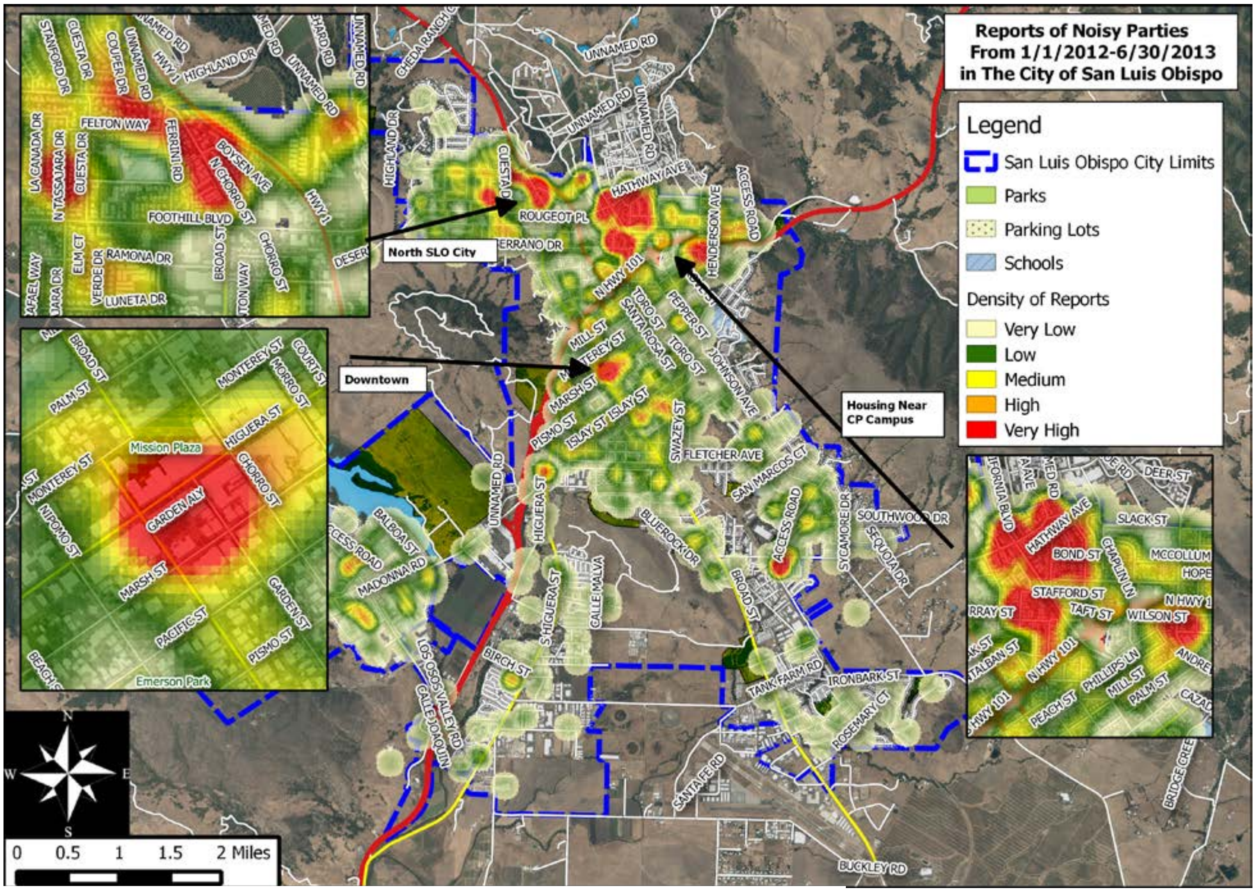


**Figure 9.** Map of all reports of noisy parties in The City of San Luis Obispo.

In this map of crime reports with a type of 'Noise Party' areas of very high density represent areas with a density of 0.13 reports within a 500 foot radius. This map clearly shows that reports of 'Noise Party' are concentrated in the downtown area and at the north end of the city near the university campus. The largest hotspot is centered on the intersection of Stafford Street and California Boulevard and extends about a quarter of a mile in each direction. This area of highly composed of rental properties typically rented by students. This is similar of the area at the north end of the city as well near Highland Drive. The hotspot in the downtown area is centered on the restaurant and popular music venue SLO Brewing Company. These represent areas of high concentrations of students and in the case of the downtown area a popular music venue with weekly concerts.

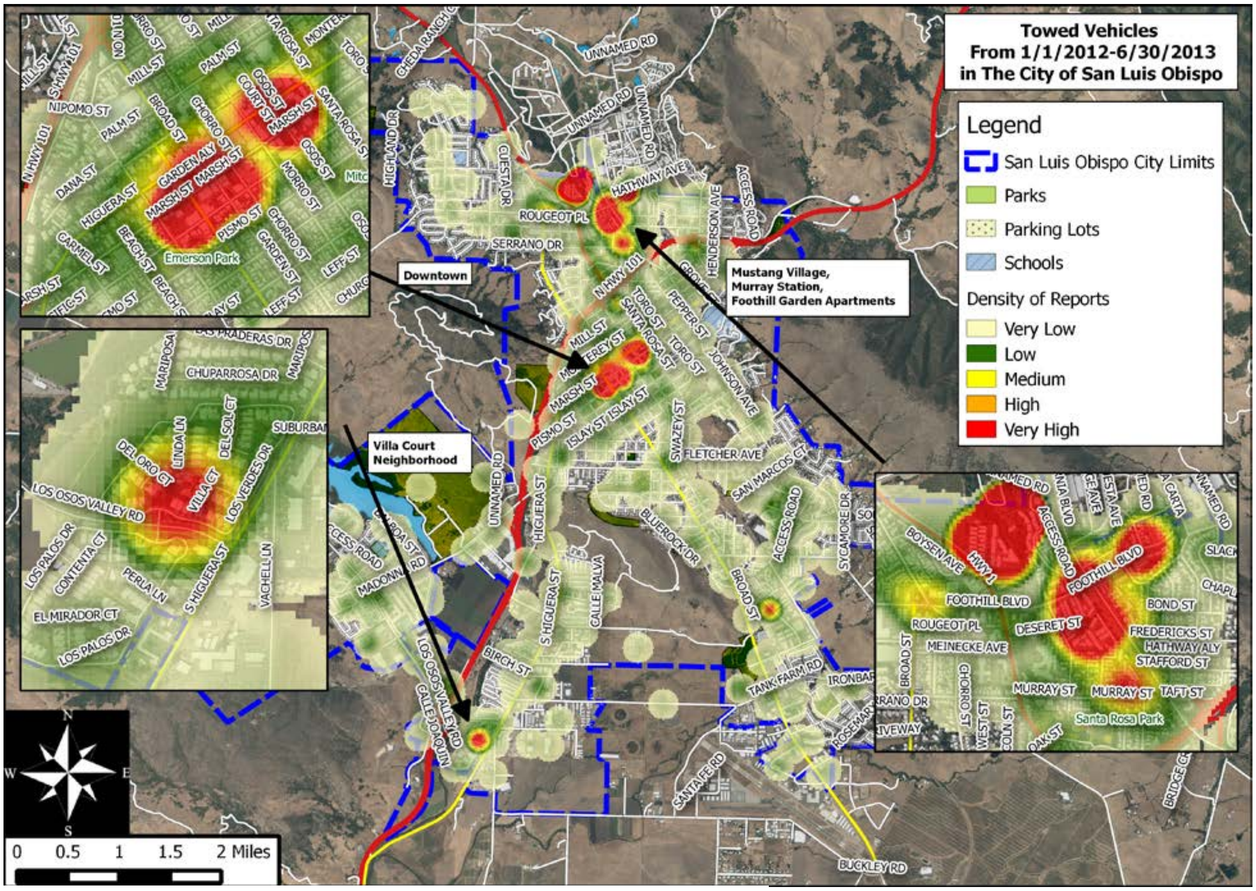


Figure 10. Map of all towed vehicles in The City of San Luis Obispo.

This map of all towed vehicles between January 2012 and June 2013 showed some of the clearest trends in my dataset. Red areas in this map represent areas with a density of 0.21 reports of towed vehicles in a 500 foot radius. The largest hotspot occurred at the north end of the city covering Mustang Village and Murray Station Apartments, both large apartment complexes housing mostly students. Mustang Village reported the most towed vehicles of any address with 238 vehicles towed in an 18 month period. This was the most common type of report for an address with recurring crime reports. The downtown area showed the second largest hotspot which can be attributed to the high concentration of drivers and also the city funded parking program which employs several parking officers.

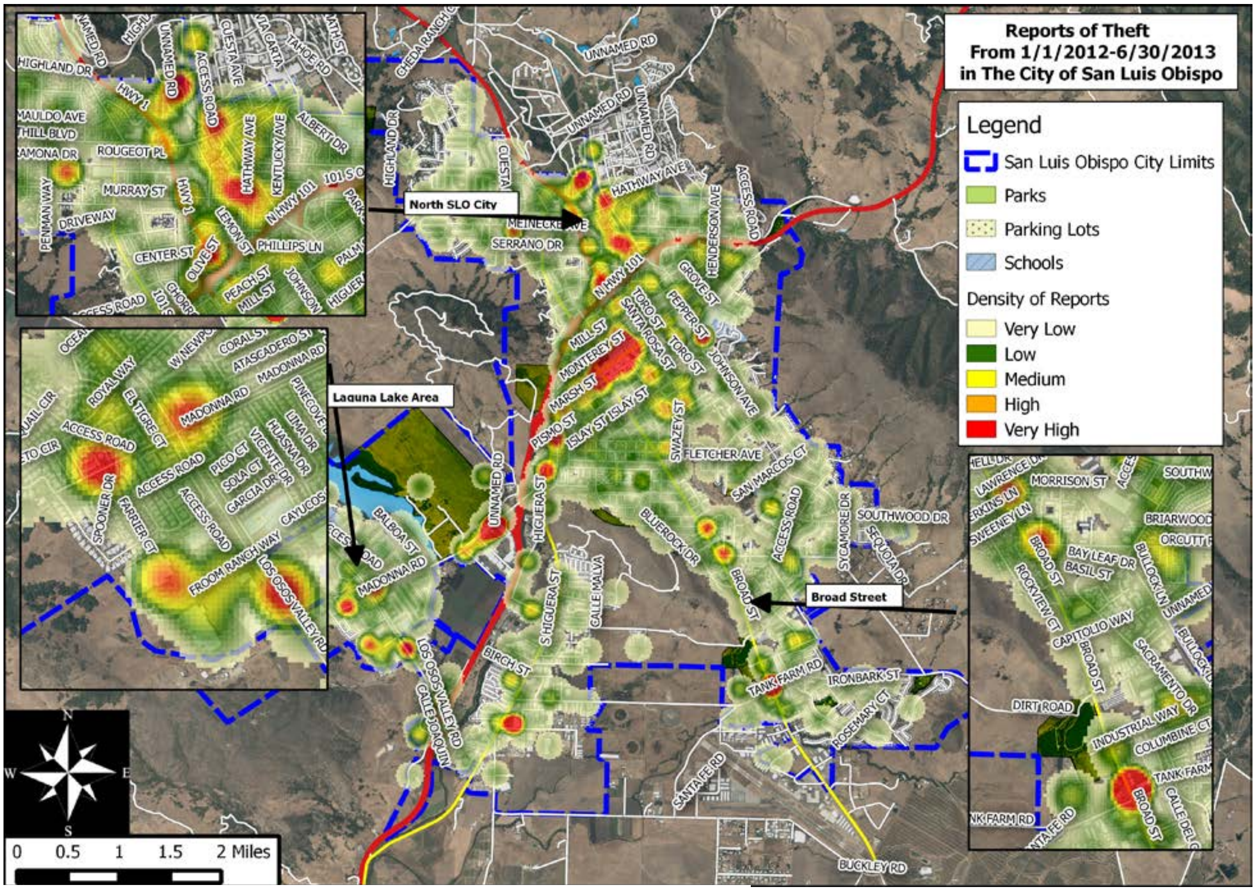


**Figure 11.** Map of all reports of theft in The City of San Luis Obispo.

This map showing all reports of theft within the city shows many scattered hotspots with the downtown area being the largest. The red zones on this map represent areas that had a density of 0.23 reports of theft in a 500 foot radius. Besides the hotspot centered on the downtown area there are hotspots occurring at the north end of the city near the areas of student housing, in the neighborhoods near Laguna Lake, and on south Broad Street. The downtown area is home to many retail stores which are constant victims of theft which attributes to the hotspot there. Clear reasons for the hotspots for the Laguna Lake area and the area at the north end of the city are few other than they are residential areas where theft also tends to occur. The hotspots along the south section of Broad Street all occur at areas where shopping centers are located.

These maps serve to show where the specific types of crime generally occurred for the period of my dataset and help to show the buildings and streets where the hotspots occur. I was also able to generate a list of the most common addresses that reported crimes:

| Number | Address | Count | CPN | Type |
|---|---|---|---|---|
| 1 | 1 MUSTANG | 446 | Mustang Village | Student Housing |
| 2 | 725 HIGUERA | 261 | Mother's Tavern | Bar |
| 3 | 1400 OSOS | 258 | Mitchell Park | Park |
| 4 | 1010 MURRAY | 253 | Sierra Vista Regional Medical Center | Hospital |
| 5 | 670 HIGUERA | 238 | West End Espresso & Tea | South end of creek |
| 6 | 765 FOOTHILL | 230 | Rite Aid | Drugstore |
| 7 | 3860 HIGUERA S | 215 | WSImagicweb, Digital Pop Marketing | Business building |
| 8 | 158 HIGUERA | 215 | Higuera & South | Business building, Gas station, restaurant |
| 9 | 989 CHORRO | 210 | The Network Shopping Center | Creekside |
| 10 | 1911 JOHNSON | 193 | French Medical Center | Hospital |
| 11 | 692 MARSH | 190 | 7-Eleven | Convenience Store |
| 12 | 1540 FROOM RANCH | 184 | Costco | Store |
| 13 | 990 PALM | 180 | City Hall; Downtown Transit Center | Bus stop |
| 14 | 3180 BROAD | 165 | Chevron | Gas station |
| 15 | 190 SANTA ROSA | 162 | Santa Rosa Park | Park |
| 16 | 3900 BROAD | 158 | Autozone | Autoparts Store |
| 17 | 1625 CALLE JOAQUIN | 158 | Motel 6 | Hotel |
| 18 | 43 PRADO | 156 | Prado Day Center | Homeless shelter |
| 19 | 200 SANTA ROSA N | 156 | Mustang Village | Student Housing |
| 20 | 11990 LOS OSOS VALLEY | 154 | Target | Store |

**Figure 13.** List of 20 most common repeat addresses.

This list shows the top 20 repeat addresses that occurred throughout my dataset and there are clear similarities within the group. The list leader is the student housing complex Mustang Village which also comes in at the 19th spot at its second address. The list has both major hospitals in the city on it, Sierra Vista Regional Medical Center and French Hospital as well. Stores downtown that border the creek also reported high amounts of crime in the dataset with a combined 448 crimes reported. The list also has the addresses of several business buildings, a homeless shelter, and retail stores. There are no single-family residences represented in the list which furthers the idea that more crime is reported in areas of high population density.

Along with the maps based off of spatial data I was able to create maps showing the dates with the highest amounts of reported crimes. The top 10 highest dates for crime tended to be on the weekend and also tended to occur on or around the date of a holiday or a significant date for Cal Poly such as graduation or the Week of Welcome.

| Date | Day of Week | Date Significance | Crimes Reported | Top Types of Crime |
|---|---|---|---|---|
| 9/14/2012 | Friday | Cal Poly WOW | 153 | Suspicious, Alcohol Offense |
| 3/17/2013 | Sunday | St. Patrick's Day | 129 | Alcohol Offense, Disorderly |
| 10/27/2012 | Saturday | Weekend before Halloween | 123 | Alcohol Offense, Noisy Party |
| 9/22/2012 | Saturday | Cal Poly 1st Week of Classes | 122 | Noisy Party, Alcohol Offense |
| 6/24/2013 | Monday | NONE | 119 | Alarm Audible, Suspicious |
| 10/26/2012 | Friday | Weekend before Halloween | 118 | Alcohol Offense, Theft |
| 3/30/2012 | Friday | Weekend of Cesar Chavez Day | 116 | Theft, Noisy Party |
| 6/15/2013 | Saturday | Cal Pol Graduation | 114 | Noisy Party, Alcohol Offense |
| 10/5/2012 | Friday | NONE | 114 | Towed Vehicle, Noisy Party |
| 9/15/2012 | Saturday | Cal Poly WOW | 114 | Noisy Party, Alcohol Offense |

**Figure 14.** List of 10 dates with highest amounts of reported crimes and their most common type.

As shown in the table there are four dates significant to Cal Poly on the list including both the Friday and Saturday of the Week of Welcome in 2012, the first Saturday in Cal Poly's Fall quarter of 2012, and the Saturday of Cal Poly's graduation ceremonies in 2013. Other significant dates on this list are related to holidays including Halloween, St. Patrick's Day and César Chávez Day in 2012. The remaining two dates have no obvious correlation with a holiday or a significant Cal Poly date yet still experienced similar amounts of crime as the rest of the dates in the list. This list was quickly generated using a free plugin for QGIS known as GroupStats that allows users to create tables similar to Excel's pivot tables. To better examine these high crime dates I created heatmaps based off of their point density which can be seen below.

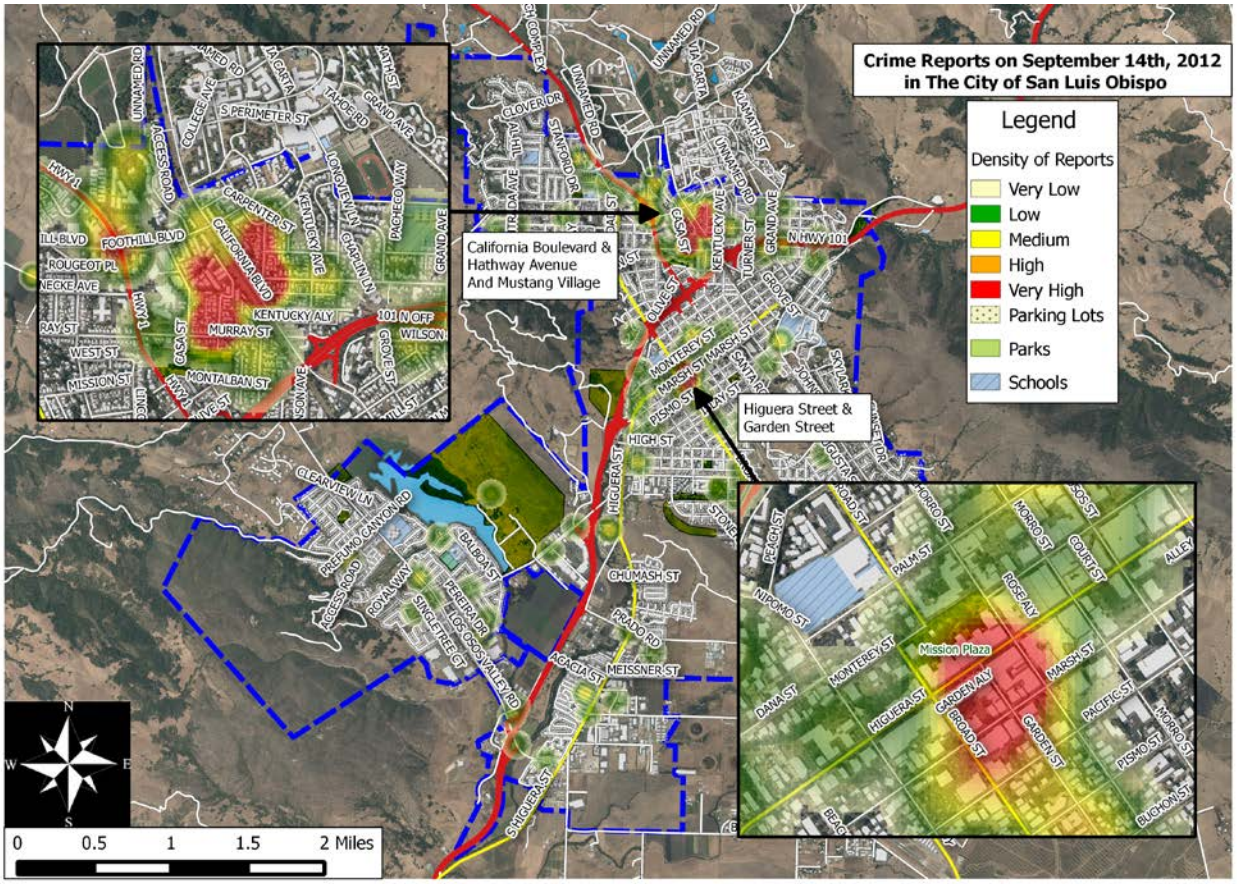


**Figure 15.** Reports of crime on September 22, 2012.

The above map shows the distribution of crime reports on September 14th, 2012 which was a Friday during that year's Week of Welcome put on by Cal Poly San Luis Obispo for incoming freshmen students. Crime reports on this date were most dense in the student housing areas near campus of Mustang Village, Murray Station Apartments, and along California Boulevard between Hathway Avenue and Foothill Boulevard. The most common type of crime report on this date was 'Suspicious Activity' and 'Alcohol Offense'.

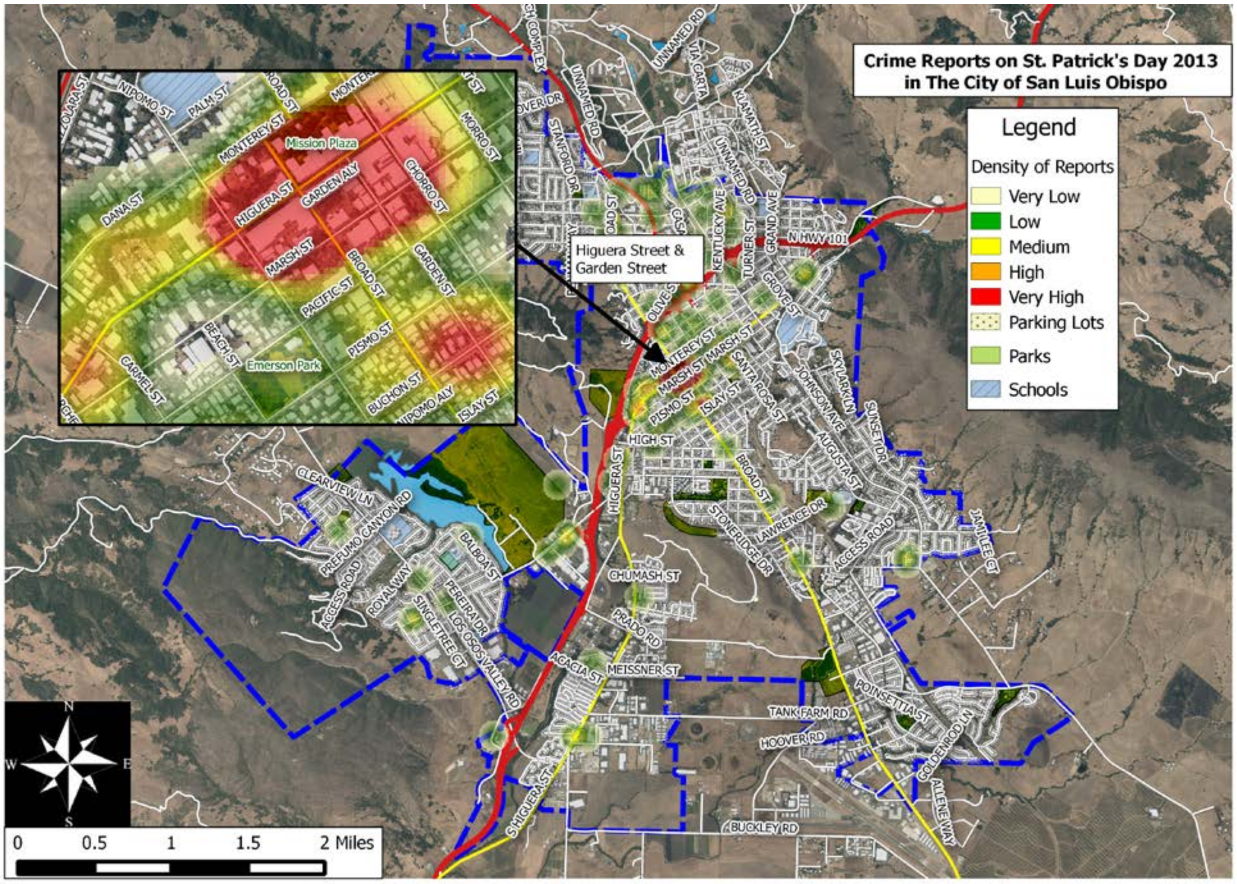

**Figure 16.** Reports on March 17th, 2012 or St. Patrick's Day.

This map details the density of crime reports on March 17th, 2012 or St. Patrick's Day. Crime reports on this day were focused in downtown San Luis Obispo around the Mission Plaza area on Higuera Street. The most common type of crime on this day was 'Alcohol Offense'. The high concentration of 'Alcohol Offense' reports in the downtown area can be attributed to the high concentration of bars and restaurants. Also on St. Patrick's Day, the San Luis Obispo Police Department makes a concerted effort to monitor crime which can be another reason as to why this day was a high crime day.

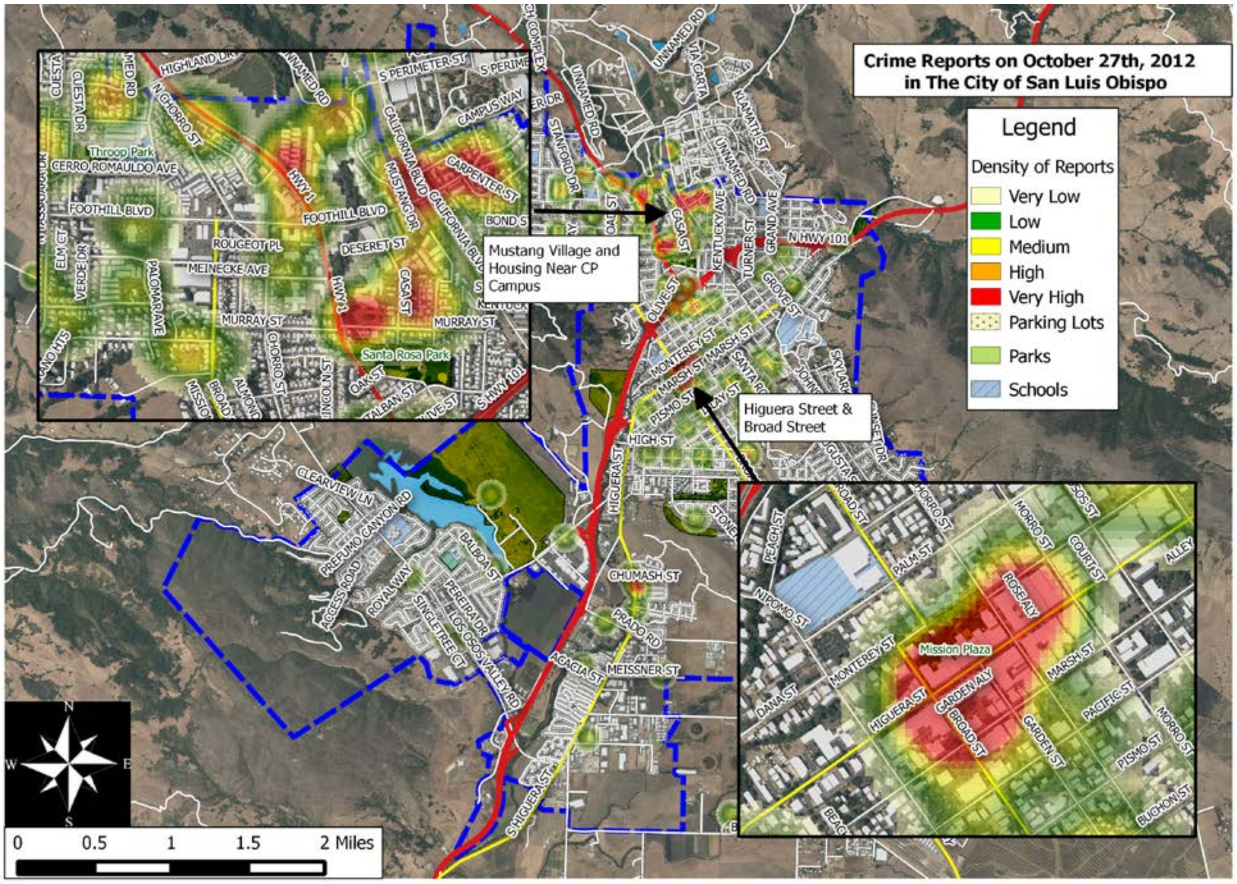


**Figure 17.** This map shows the concentration of crime reports on October 27th, 2012.

This map showing the concentrations of crime reports on the Saturday before Halloween in 2012 highlights a few obvious hotspots. One hotspot occurs downtown at the intersection of Higuera Street and Broad Street where several Bars are located. Another group of hotspots exists at the north end of the city near Cal Poly campus. The most common crime reports on this day were 'Alcohol Offense' and 'Noisy Party' which can be attributed to the celebrations surrounding Halloween.

**Conclusion**

Clear trends emerge after examining the data based on a myriad of factors which help to define areas and specific times of the year when crime is being reported. It is obvious from the data that the majority of crimes reported in the City of San Luis Obispo originated from the north part of the city near the university and from the downtown area. Other areas of high reported crime emerged as well along south Broad Street and the Madonna Shopping Plaza. Dates of high crime tended to coincide with either significant dates related to the university or popular holidays like Halloween and St. Patrick's Day. My report proves that complex crime analyses can be performed with FOSSGIS and can produce real results that identify areas of high and low density crime. Although I was able to perform all of my analyses with free software, I experienced difficulty at some points in the process. Converting the police department's text file of all the crime reports into a usable CSV took up a significant portion of my time. Luckily I was able to receive help from the online FOSSGIS community as there is no customer support for QGIS outside of that. Another bottleneck in my analyses was that I was unable to run QGIS on my home computer because its operating system is too old. In order to work around this I did all my work at the Robert E. Kennedy Library on the Cal Poly Campus with QGIS loaded onto an external hard drive. I ran QGIS from my external hard drive because users are not granted administration rights on the computers in the library which means programs cannot be downloaded to the machines there. It is my hope that QGIS be eventually loaded onto the GIS computers in the library to expose students to an alternative GIS software package. Besides these two main restrictions I was also limited on time I could devote to my crime analysis. In the future I plan to produce a shareable interactive webmap via Mapbox that showcases my findings and also a printed map showing the main hotspots for crime in the City of San Luis Obispo.

# Works Cited

1. Santos, Rachel Boba. *Crime Analysis and Crime Mapping*. Thousand Oaks, CA: Sage Publications, 2005. Print.

2. Hart, Timothy C., and Paul A. Zandbergen. "Reference Data and Geocoding Quality Examining Completeness and Positional Accuracy of Street Geocoded Crime Incidents." *Policing* 36.2 (2013): 263-94. Print.

3. Harvey, Francis, Nicholas R. Chrisman, Andrew Frank, Dan Sui, Waldo Tolber, and Michael F. Goodchild. *Are There Fundamental Principles in Geographic Information Science?* S.l.: Association of American Geographers Geographic Information Systems and Science Specialty Group, 2012. Print.

4. Steiniger, Stefan, and Andrew J.S. Hunter. "The 2012 Free and Open Source GIS Software Map - a Guide to Facilitate Research, Development, and Adoption." *Computers, Environment & Urban Systems* 39 (2013): 136-50. Print.

5. Ballatore, Andrea, Michela Bertolotto, and David C. WIlson. "Geographic Knowledge Extraction and Semantic Similarity in OpenStreetMap." *Knowledge and Information Systems* 37.1 (2013): 61-82. Print.