Spring 5-5-2015

Map Centered Analysis of San Francisco Crime Trends

Brittany Moore
Syracuse University

Follow this and additional works at: https://surface.syr.edu/honors_capstone

Part of the Technology and Innovation Commons

Recommended Citation
https://surface.syr.edu/honors_capstone/879

This Honors Capstone Project is brought to you for free and open access by the Syracuse University Honors Program Capstone Projects at SURFACE. It has been accepted for inclusion in Syracuse University Honors Program Capstone Projects by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.
A Map-Centered Analysis of San Francisco Crime Trends

A Capstone Project Submitted in Partial Fulfillment of the Requirements of the Renée Crown University Honors Program at Syracuse University

Brittany Moore  
Candidate for Bachelors of Science Degree  
and Renée Crown University Honors  
May 2015

Honors Capstone Project in Information Technology

Capstone Project Advisor: _______________________________.  
Interim Dean of the School of Information Management Jeffery Stanton Ph D

Capstone Project Reader: Professor of Practice Jeffery Saltz

Honors Director:  
Stephen Kuusisto, Director

Date: April 23, 2015
Abstract

The original purpose of my Capstone was to analyze crime statistics in San Francisco to draw conclusions that could help prevent crime. I chose to analyze crime statistics in my hometown because I believe that using a place that I’m familiar with and engaging in a project that would be addressing a problem that has affected my family would help to motivate me. I’m also deeply passionate about using data analysis skills in ways that will directly improve the human condition. I did this by accessing a dataset online, using R Program Language and creating maps of relationships of crimes throughout the city.
Executive Summary

In my database class, I learned about simple functions you could do that could give you a unique insight into a big database that might otherwise appear to be boring. It was always exciting to me that an amazing world-changing story could be found in a big pile of seemingly boring data. Big Data, or large amounts of data about anything, has been around for a long time, but new tools have emerged that can help people make more sense out of this data. Specifically, with social media sites like Facebook, Twitter and Instagram you can collect more information about human behavior, opinions and trends than ever before.

I was heavily influenced by the way information management was used in the Obama campaign. I read about how data scientists (or geeks) used information they could collect about people, and about their habits and behaviors, to best target them and help Obama win. I loved the idea of using data to create a positive outcome. Going to talks about Big Data at my school, I discovered that R programming language was based on statistical analysis and could be used to manipulate large amounts of data.

Big Data could also be used in several areas. Specifically, in regard to geographic data, specific information could be given to real estate companies for a greater understanding of specific locations. It is easy to see that this information could easily be misused to further red-line disadvantaged communities or to oversimplify crime-ridden areas, without looking deeper to find stories that explain these crimes (like lack of employment because of racism in hiring processes). Many companies, like Facebook, sell bundles of their data to advertisers to gain a deeper, psychological understanding of their target demographic to better sell to them.

The San Francisco Police Department is cutting edge in releasing all of its data to the public. Data analytics has traditionally been used for advancing the private sector. Adversities and marketers are keen on targeting people and understanding their habits to sell them products or get them to vote for a candidate. For my Capstone, I wanted to use this technology for the non-profit government sector. Using tools to better understand crime and to discover if certain kinds of crime are related will improve the effectiveness of local law enforcement and the safety of citizens and will eventually help to prevent some crimes before they happen.
# Table of Contents

Abstract........................................................................................................ iv  
Executive Summary...................................................................................... v  
Preface......................................................................................................... vii  
Acknowledgements (Optional)................................................................. viii  

Chapter 1: The Data.................................................................................... 1  
   Shifting Focus......................................................................................... 1  
   Pick a variable to focus on................................................................. 2  

Chapter 2: Charts..................................................................................... 4  
   Charts.................................................................................................. 4  
   Other Charts....................................................................................... 9  

Chapter 3: Maps ..................................................................................... 10  
   Relational Maps.................................................................................. 13  

Chapter 4: Key Insights .......................................................................... 14  
   Nature of Exploratory Independent Research...................................... 18  
   Use of Maps........................................................................................ 18  
   Mapping and Decision Making......................................................... 19  

Chapter 5: Charts vs. Maps .................................................................... 21  
   Finding Relationships......................................................................... 22  

Chapter 6: Beyond the Capstone.............................................................. 24  
   Demographic Research....................................................................... 24  
   IBM..................................................................................................... 25  

Works Cited............................................................................................... 26  
Appendices................................................................................................. 27
Preface

Growing up in Menlo Park, a city on the outskirts of San Francisco, I was extremely sheltered. Because of my parent’s protection, I was lucky enough to think the world was a safe, friendly place for longer than many children in my area did. Only when I moved to a suburb in Colorado in sixth grade did I come to realize how different my hometown was. Growing up, I didn’t know there was such a thing as delivery pizza. I didn’t realize that my neighborhood was deemed too dangerous for pizza delivery because of past events involving gun violence and murder.

Tamera

Years later, I was devastated a when a childhood friend lost her life. Tamera was a girl who lived a few blocks from me. She didn’t like her home life and would often run away for days at a time. I later found out that as a very young teenager, she had become involved in prostitution. She was in sex trafficking in the bay area for only a few years, until her body was discovered in a local park after she had been killed by a drug lord.

As painful as Tamera’s story is, I don’t think it is uncommon. In every major city, I believe, there are girls with similar stories who run away from home often and in a few years time are involved in criminal activities and, more than likely, sex crimes. If police look for runaways/missing person reports, if they fit the demographic for prostitution, young teenage girls then could be found in areas where prostitution is common, and hopefully be helped before they end up like Tamera. This foresight could influence community organizations to do programming there, or at the least provide answers to the families of these missing people.

Coming to college, I had no idea what the School of Information Management (iSchool) was or how I ended up there. But I knew I always loved learning and devouring information in the form of old books or vintage magazines. I was always curious to learn what was happening. I was always excited by the possibility of ingesting information. The more I learned about the iSchool, the more excited I became about being able to take control of much more data than I could ever read at one time.

Through college sociology classes I became more aware of larger social structures, like systemic racism in the housing market and the privations of prisons, causing dramatic inequalities, I see this exemplified in the differences between my community of Menlo Park and the community of Atherton, only a few miles away.

Having this greater awareness and personal experience of losing a childhood friend, I became curious to understand how I could use what I’ve learned to help disadvantaged communities. I was amazed at everything I was learning in school, because it answered many of the questions I’d had growing up, like why our neighborhood couldn’t receive delivery when my classmates’ neighborhoods could. With the tools learned of in the iSchool, I saw how my Capstone could be an opportunity to use the resources at my disposal to give back to my community and, hopefully, prevent others from ending up as I’d seen my friends end up.
Acknowledgements

I’d like to say a special thanks to my Capstone Advisor, Jeffery Stanton, without whose patience and guidance my Capstone couldn’t have been completed.

I dedicate the work of my Capstone to all those who have been the victims of crimes in the Bay Area, with the sincere hope that technology will be used to improve the lives of those without the privileges of others.
Chapter 1
The Data

Through the help of librarians, I discovered several government websites where public crime data was available to be downloaded in Excel form. I ended up getting my data from data.sfgov.org. This data included all documented crimes from the San Francisco Police Department from March to May of 2015. The information had 12 rows, including the categories crime (i.e., assault or larceny), description (i.e., battery or grand theft auto), date of the week, time, district and address of the crime. I was surprised at how easy it was to find and download all this information for free. There were 32,252 instances.

Shifting Focus

When writing the Capstone proposal, I wanted to get basic information about a vast amount of crimes. For example, in my proposal, I stated, “I plan to gather six histograms, box and whisker graphs of the six main categories. This will give me at least 36 information graphics to analyze (not including more graphs that I may run as combinations of the different categories). I also planned to use Excel to create charts and comparisons between crime categories.

After running my plan past my Capstone advisor, I realized I needed to shift my focus from broad to narrow. Dean Stanton encouraged me to focus solely on R and not Excel. Instead
of looking at multiple variables, I decided to dig deeper into one type of crime and learn what I could from that. We broke down the work into the following steps:

1. Get the dataset
2. Export the dataset into a comma separated values file
3. Pick a variable to focus on
4. Make some univariable stats
5. Learn about creating maps in R
6. Make a heat map of the crime stats
7. Draw conclusions.

Refocusing to do less general work and delving into only one subject enabled me to push myself further than I could have doing several less challenging tasks.

*Pick a Variable to Focus On*

During my journey in R programming, I discovered how to create a basic summary of all of my data. This little bit of code was extremely helpful for getting a quick overview and understanding the scope of the data I was working with. I was able to see the statistical breakdown of incidents in quadrants, in a numeric box and in a whisker graph. I was later disappointed to see how easy it was to understand this information with the summary feature, in comparison to trying to represent the same information in a graph. With this discovery, I decided not to do additional whisker graphs because this information was provided in the summary.
With this knowledge, I chose to use the variable of *theft and larceny* because thefts and larcenies occur the most often, and they are also diverse, in that they take place in different districts, on different days of the week and at different times.
Chapter 2

Charts

Originally, I planned to make five univariable charts on larceny/theft, to get a better understanding of the diversity of occurrences. I wanted to make bar charts of Time, Day of the Week, PD Distinctions and Descriptions, meaning what the type of larceny was: grand theft auto, petty theft or attempted theft. Throughout my research, I discovered the difficulty of accomplishing this task. I was only able to produce one chart because of the many challenges I faced in this process.

Challenges

Because R is a statistically based programming language, I had a lot of trouble trying to convert letters into numerical representations or understand words or phrases. Many of these challenges I never overcame; instead, I decided to focus on more achievable goals. For example, I wanted to create a graph that separated crimes by the day of the week. I was curious to see if a certain type of crime was more likely to happen on a certain day of the week. I ultimately wasn’t able to accomplish this. The actual days of the week, Monday, Tuesday, etc., aren’t recognized in R. Monday is seen as a numerical representation, for instance, 1, Tuesday would be represented as 2, Wednesday as 3 and so forth. Therefore, in my code, I would have had to create a key of some sort, one which first converted all of the larceny reports of a day into
numerals. This step would illustrate the number of occurrences over a numeric X value, in this case, 1-7 for the days of the week. Then in the code, I would have had to create an alteration to display the numbers as their corresponding day of the week. For example, I would tell the code whenever it sees a “1,” it should convert to the word “Monday.” Although I could map out what this process would entail, it was very difficult to find the obscure pieces of code to perform all these steps. Consequently, I decided to not do this univariable stat because it was so far off of the basic uses of R. As mentioned, R is a statistics-based computer code. It works best with numbers, not words, and it was very difficult to change all of the words categorizing the crimes to numbers.

I encountered a similar struggle creating a chart of the times when crimes occurred, but I was able to create a basic chart. The statistical nature of R made it challenging to create a time-based bar chart as well. The times when crimes occurred were entered into the dataset in a different format than I expected. I was unsure if it was by military time or by the number of minutes in a day. Ultimately, I decided not to pursue creating a map broken down by time. The package that reads time was decoding the time in minutes instead of hours. I struggled again with finding how out to reformat this into hours or times of day (morning, afternoon, evening, etc.). With the chart I was able to make, I could see that there was a peak in larceny during the 10-11am ranges.
Figure 2 {Basic Histogram}

I was able to make the most progress with the Police Department District Bar Chart. Thankfully, because the PD District names are all unique, I could relatively easily get the names on the $x$ axis. The largest problem I ran up against was getting the information to be easily readily and comprehensible. I researched several ways to experiment with the $x$ axis. First, I discovered the chunk of code that accurately spaces the data, in this case the 10 PD Districts, so that it occupies the entire $x$ axis by unitary increments. Also, I discovered how to tilt the names on the axis by degrees to make them legible. In addition, I added a $y$ axis label and title, and turned the bars pink. Below is my evolution.
Although Figure 3 may look simple, it took hours to achieve this. The biggest problem was getting the correct amount of spacing on the $x$-axis so the names would not overlap with each other. Then the focus was on rotating the names (down to the degree) to a place where they could easily be seen. Figure 4 shows the code that was needed to create Figure 3.

```
> barplot(table(larcency_theft_stats.csv$PdDistrict), main="PD District", xlab="", ylab="Number of Crimes", space=0, axisnames=FALSE, col="pink")
> text(x=seq(1,10,by=1),srt=-45,adj=0,xpd = TRUE,labels=rownames(table(larcency_theft_stats.csv$PdDistrict)),cex=0.65)
```

Figure 3 {PD District Graph}
Figures 5 and 6 illustrate the process of achieving a readable graph.

Initially, I enjoyed the fact that R was very command-line based, but problems like these show how a clunky, command-line-based program makes doing simple tasks very difficult, although it is possible to figure out how to do them with patience and perseverance. If R had a simple GUI that was user friendly, I could easily have created graphs with a few clicks, instead of spending hours working on the labels and spacing. Struggling with this graph gave me a new appreciation of desktop-friendly applications and an understanding of how much work goes into making software user-friendly.
**Other Challenges**

One of my biggest challenges was the process of organizing all of my work. I found myself being stuck for a while and then making a series of breakthroughs, which then led me to the next sticking point. The process of organizing all of the material was very time consuming. I had to work out a perfect balance of not retaining code that turned out to be incorrect, versus forgetting to save code that turned out to be correct. My research was exploratory and filled with trial and error. I eventually organized all my work into a master script, a series of code for the command-line, which can be reproduced.

I was most surprised by the granular manipulability of the visuals, although it was a challenge finding the specific bits of code in textbooks or detailed Google searches for minor problems I was having. With more time and searching, I could have found a more detailed map of San Francisco, with district and landmark outlines. I also could have changed the style, shape, size, size of the dots on the map and so forth. All of the products I used for this research, including the textbook, software, and every piece of data, were available for free online. With more time, expertise and focus, I could have created an extremely complex and detailed presentation.
Once I began the process of creating a map of my findings, I originally I plotted all the points of crime I was given in my spreadsheet. Figure 7 shows all 32,252 instances that I was given. It’s clear to see that there is a concentration in the southern Bay area. I used a very basic outline of SF that I could find online in the shape file format (as this specific format is needed to be compatible with the R programming software I was using). Figure 8 is a more detailed image of SF, showing more boundaries and landmarks, which can be used for comparison.

Figure 7 {All crimes mapped}
Then I began to filter my results by different types of crime.
Figure 9 {Occurrences of Larceny/Theft}

Figure 10 {Drug/Narcotics crimes}
Looking at only one variable at a time gave me a good idea of the crimes’ geographic layout. Next, I was interested in seeing the similarities among different categories. I was interested in seeing if there was any correlation between different kinds of crime. In R, I began laying maps over each other.
Figure 14 shows the rape stats in green and missing persons in purple. I was interested in seeing if the location or occurrences of missing people had any relation to rape victims. I thought maybe there is a correlation between people running away and being raped, or maybe those who are raped are then likely to run away. My results will be shared in the Key Insights section.
Figure 14 {Rapes in Green and Missing Persons in Purple}

Figure 15 {Larceny in Black and Vehicle Theft in Blue}
Throughout my research, I found that more information could be found by relating certain crimes to others. Through combining crime instances, I think there are more interesting stories that can be told. I drew on my own knowledge of the life of my friend, who was often a runaway, to see relationships between missing persons and rapes. I wasn’t able to find a clear relationship, because there are several other factors that can result in rapes and missing person’s reports. I think with more information on the demographics and backgrounds of these people, more correlations could be found.

I see a clear relationship between larceny and theft. Under common law, larceny is defined as “the trespassory taking (caption) and carrying away (removal) of the tangible personal property of another with the intent to deprive him or her of its possession permanently.” Specifically, around the blocks of California Street in Nob Hill and Chinatown, there is a high amount of both larceny and vehicle theft. Looking at vehicle theft alone, one can see that it is fairly evenly scattered around the city, but when combined with larceny (a smaller form of theft), it is clear that there are a few highly concentrated blocks of the city where the two happen almost within feet of each other.

---

1 California Theft / Larceny
I also found a relationship between rape and drug/narcotic crimes. It is obvious that there are a lot of crimes in the Chinatown area, both rape and narcotics crime. There are also both rape and drug crimes, although significantly fewer, in the Mission District. However, the two are not always related. I was curious to find that there aren’t as many rapes at the Japan Town and Pacific Heights border, although there was drug activity reported. This can be seen in Figures 10, 12 and 13. It would be interesting to look for more differences between Japan Town/Pacific Heights and places like the Mission and Chinatown. I focused on Chinatown, which is a known crime-ridden area, which can be seen down to the street level.

Figure16 {Magnified view of Chinatown larceny in black & drugs in orange}
**Nature of Exploratory Independent Research**

The nature of my research was exploratory. I had no set conclusion or hypothesis that I was looking for, and, therefore, was difficult to know when I had finished. I decided to explore some of the research that has been done on data mapping and decision making. Comparing past research to what I recently did, I was happy to see that I saw similarities. Much of my personal research supports the conclusions other researchers have made.

**The Use of Maps**

The University of Idaho’s 2010 report *Map-Centered Exploratory Approach to Multiple Criteria Spatial Decision-Making* explored many of the issues I looked at. Specifically, the paper covered what the use of maps should be. Of course, looking at maps or visual data is not appropriate solution for every situation.

The report states that “[m]aps become a ‘visual index’ through which the user orders decision options, assigns priorities to decision criteria, and augments the criterion outcome space by map-derived heuristic knowledge”2 (Jankowski, 4)” Yet the tools for creating high quality maps are still relatively new and are not advanced enough to enable spatial decision-making calls. It would be difficult to completely understand the spatial makeup of a region. I saw this parallel in my own work, with my not being able to make find a detailed map of San Francisco that I could use to plot my points. It is disappointing that researches are not able to make more advanced and relevant maps.

---

2 *Map-Centered Exploratory Approach to Multiple Criteria Spatial Decision-Making.*
In terms of use, the research indicated that maps are used more for analysis in the aftermath of events instead of for initial exploratory tasks. “For experts, a map was mainly a convenient tool used to detect discrepancies between the results obtained from a model and experts’ expectations based on some implicit preferences.”2 My exploratory process consisted of reflective analysis driven by a specific interest (rapes, prostitution, etc). Reflecting on my experience, I question how useful maps will be in predictive analysis for regions that are not possessed of an abundance of past data collection.

**Mapping and Decision Making**

The University of Georgia’s research paper entitled *Using Geographical Information Systems for Decision Making: Extending Cognitive Fit Theory to Map-Based Presentations* dug into issues of using maps for decision making. Specifically, the paper was based on “Cognitive Fit Theory (CFT) [which] was developed to explain how graphical displays affect the decision processes and outcomes of decision making. … The outcomes of decision making depend upon the fit between information presentation, task, and decision processes used by the decision maker” (Dennis 3).3 This idea is extremely important for law enforcement’s making quick decisions, especially in the heat of the moment, as they often must do.

The paper then explored what occurs when the data does not show what someone would think. “When a mismatch occurs, one of two processes will occur. First, decision makers may transform the presented data to better match the task, which might increase the time needed and might decrease accuracy because any transformation can introduce errors3 (Vessey 1991). Alternatively, decision makers may adjust their decision processes to match the presentation.

3 *Using Geographical Information Systems for Decision Making: Extending Cognitive Fit Theory to Map-Based Presentations*
(Perrig and Kintsch 1985), decreasing accuracy and increasing the time, because the information does not match the ultimate needs of the task. Someone’s reaction to this data and decision mismatch would be a dangerous error. The very nature of making tough decisions means that you do not have all of the information, and your decision may turn out to have been wrong. Even with data analysis tools, these decisions will remain tough for law enforcement and may be more impactful for communities.
Chapter 5
Charts vs. Maps

“CFT classifies decision making processes used for the elementary tasks as being analytical or perceptual (Vessey 1991). Analytical processes emphasize precision and the processing of information based on its Style of Presentation” (Dennis, 7). Charts are best used for analytical purposes, and maps are perceptual. I agree with the hypothesis in the paper that “[m]ap-based based presentations will induce perceptual decision processes, while tabular presentations will induce analytical decision processes” (Dennis, 5).

Getting information into an Excel spreadsheet is not going to help law enforcement, who deal with the city and its people, not numbers and percentages. “Spatial presentation should result in faster performance because it induces perceptual processes which are inherently faster than analytical processes” (Vessey 1994; Dennis, 5). They found, upon conclusion of the study, that it was clear that “[d]ecision makers using the map-based presentation were more likely to use perceptual rather than analytical decision processes” (Dennis, 8). It is clear that for people who have to make immediate decisions and are held to the standard of protecting communities, maps and further data visualization tools produce more accurate and faster results for multi-criteria data. Hopefully, as this knowledge spreads, mapping tools will become more advanced to help with decision-making.
**Finding Relationships**

A similar concept from both readings is that maps are best for analyzing more than one element, and thus finding relationships. Maps are best when looking at specific “information-seeking tasks” and best with multiple criteria and other data analysis tasks. “Malczewski (1999b) suggests that the main objective of using maps in multiple criteria spatial decision analysis should be the consideration of geographical locations in the process of exploring trade offs among the decision criteria and the search for best (compromise) solutions to the decision problem” (Jankowski, 6).

“The experiment found that decision makers using a map-based presentation made faster and more accurate decisions when working on a geographic task in which there were adjacency relationships among the geographic areas. Decision makers using a map-based presentation made faster but less accurate decisions when working on a geographic task in which there were no relationships among the geographic areas” (Dennis, 10). Through my research, I would agree that maps are best when working on relationships—that’s what data mining and map tools are best for—especially with human behavior data, everything is relational.

Dennis’ research found that “[f]or (a) task, which required an understanding of the relationships among the geographic areas, subjects using map-based presentations made more accurate decisions in less time” (Dennis, 8). In the process of creating maps versus graphs, I found it quicker to identity issues with a map. “Our conclusions are that map-based presentations help decision makers in some situations. In cases where it is important to understand the relationships among geographic areas, map-based presentations improve performance. Using
map-based presentations on geographic data for which an understanding of relationships is not needed (i.e., a geographic containment task) is more problematic” (Dennis, 10).
Chapter 6
Beyond the Capstone

If I were to continue this work, I would want to continue to look at the relationships of certain crimes to others. I believe that many crimes are related or are exaggerations of each other. For example, for many of the places where assaults happened, rapes were close by, if not in the same places. Looking into predictive analytics, I predict that places with a high amount of assaults have a greater likelihood of rape. I would like to look into the locational relationship of theft and drugs. I would like to know if you can trace what stolen items are used for. Are thefts being used for drug money, and is there any data that supports that?

Demographic Research

I am interested in understanding the demographics of the people committing these crimes. I would like to break down types of larceny/theft into large crimes, medium and petty crimes. If possible, I would like to link the neighborhoods where these criminals live and their ages; this would be helpful for understanding what crimes are most associated with what socioeconomic backgrounds and ages. Using mapping data to better understand the people whom the police have the most contact with would be very useful, given the recent history of racial profiling and the killings of unarmed black men. I think there are some interesting trends that
could be found from these relationships. I predict that young people who commit petty thefts are more commonly from lower socioeconomic backgrounds and are stealing to provide for their needs. Having this data could help with the adjudication process in determining appropriate and reasonable punishments for different crimes.

I would also continue researching sex trafficking/prostitution and drugs and runaway people in major cities. I see the story of my friend Tamera as the story of thousands of girls like her in cities around the world. Lots of progress needs to be made in researching this specific issue, as prostitution was not a category in the data set I worked with. For example, in looking for runaways/missing persons, if these people fit the demographic for prostitution, young teenage girls then could be searched for in the areas where prostitution is common. This foresight could influence community organizations to do programming there, or at least help to give answers to the families of these missing people.

**IBM**

Moving forward, I have accepted a position in the Big Data division of IBM as an IT specialist. In this role I will work on cutting edge data analysis and visualization software and explore how they can best fit the needs of IBM’s global clients. I’m interested in seeing how far technology evolves with analysis and how this will shape the uses of Big Data. Hopefully, clients will use this technology in ways that benefit the millions of people who use their services. I will also keep monitoring how this technology is used in law enforcement and in the non-profit sector.
Works Cited


Appendices

Copy of R Code

```r
setwd("/Users/OtherSums/Desktop/CaseNotes")
on <- read.csv("larceny_theft_stats.csv", header = T)
lineage_theft_stats <- read.csv("larceny_theft_stats.csv", header = T)
View(larceny_theft_stats)
barplot(table(larceny_theft_stats$District),
       main = "Number of Crimes", sub = "larceny_theft_stats.csv$District",
       xlab = "District", ylab = "Number of Crimes", par.settings = par.settings,
       col = "pink")
tex(x = seq(1, 16, by = 1), y = 45, adj = 0, xpd = TRUE, labels = rownames(larceny_theft_stats$District),
cex = 0.65)
install.packages("maptools")
require("maptools")
install.packages("PBSmapping")
require("PBSmapping")
shapefile <- read("QGISdata\larceny\2013\us_cities10m")
summary(shapefile)
# (shapefile) this will delete the shapefile!!!
plotPoly(shapefile, xlim = c(122.5, 122.4), ylim = c(37.71, 37.82))

# those two lines create events for all the crime reports and then plot them to map
edffold <- as.EventData(data.frame(EID = as.numeric(rownames(larceny_theft_stats)),
                                  X = larceny_theft_stats$X, Y = larceny_theft_stats$Y))
addPoints(edffold, col = "red", cex = 0.65)

# this creates a new dataframe with just vehicle theft category
newTheftData <- larceny_theft_stats[subset(larceny_theft_stats$Category == "VEHICLE THEFT")]

# creates vehicle theft dataframe of only x ys
VehicleTheft <- as.EventData(data.frame(EID = as.numeric(rownames(newTheftData)),
                                          X = newTheftData$X, Y = newTheftData$Y))

#Making it into a heatmap
VehicleTheft <- as.EventData(data.frame(EID = as.numeric(rownames(VehicleTheft)),
                                          X = VehicleTheft$X, Y = VehicleTheft$Y))
addPoints(VehicleTheft, col = "blue", cex = 20)

# do a plot over with blank map. this takes a while
plotPoly(shapefile, xlim = c(122.5, 122.4), ylim = c(37.71, 37.82))

# another one focusing in on ChinaTown Area
plotPoly(shapefile, xlim = c(122.44, 122.4), ylim = c(37.76, 37.82))

# want do do graphs for "SEX OFFENSES, FORCIBLE", "MISSING PERSON", ASSAULT

# rape
newRAFEData <- larceny_theft_stats[subset(larceny_theft_stats$Category == "SEX OFFENSES, FORCIBLE")]
Rape <- as.EventData(data.frame(EID = as.numeric(rownames(newRAFEData)),
                                 X = newRAFEData$X, Y = newRAFEData$Y))

# Missing Person
newMissingData <- larceny_theft_stats[subset(larceny_theft_stats$Category == "MISSING PERSON")]
```