

# **Exploration Of Crime In Boston & Neighborhoods**

Tian Yu([ytian2@wpi.edu](mailto:ytian2@wpi.edu), github: ytian2)  
Naihui Wang([nwang@wpi.edu](mailto:nwang@wpi.edu), github:Naihui)  
Yang Liu([yliu15@wpi.edu](mailto:yliu15@wpi.edu), github:erouirior)

**Final Project Process Book  
CS573 (FALL 2015)**

## 1. Overview and Motivation:

As everyone know, there is a tragedy in Paris caused by ISIS on Nov, 15<sup>th</sup>, 2015, and one day after that, weapons were stolen from Worcester armory. At the same day, Harvard University even received bomb threat. Just a few days before, all cars were smashed in a parking lot on a street on which many of our WPI students live.

Under this circumstance, we decide to do this project about crime incident in Boston neighborhood. From this project, we are trying to let police know how many crimes are there in our daily lives? What kind of crime it is? When and where will crime happen mostly? We want to give useful information for Boston polices to have a rough understanding of crimes. And let them always be prepared for the crime. Even if there is only one kind of people can be more alert about crime from our visualizations, we can say we did a good job, this visualization is useful.

## 2. Related Works:

Our project is focused on crime data, thus we intently search visualization projects of crime analysis.

First, we are inspired by below figure (figure1) is from website:

<http://public.tableau.com/profile/jeffs8297#/vizhome/ACincinnatiCrimeVisualization/StartPage>. Data described in this figure are almost in common with ours. We want to also display our dataset using bar chart.

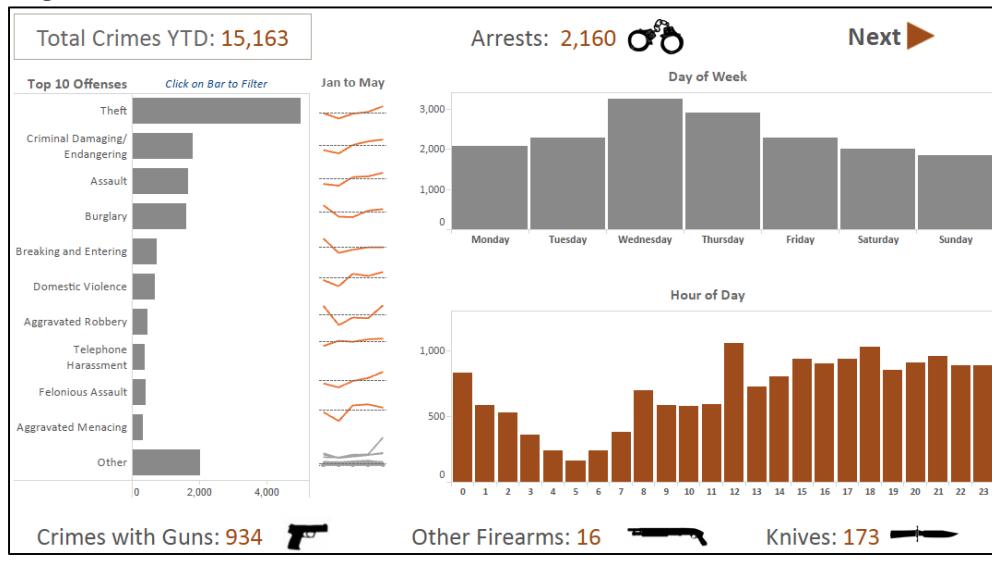


Figure1

Initially, we wanted to design a map visualization like below two figures (figure2 and figure3), figure2 is from:

<http://public.tableau.com/profile/jeffs8297#/vizhome/ACincinnatiCrimeVisualization/StartPage>; figure3 is from: <https://www.wprdc.org/showcase/city-of-pittsburgh-crime-visualizations/>. We think it is really useful to show each crime place. But in the consideration of our skill level and limited time, we gave up this map visualization.



Figure2

In below figure, we think the designer really display data in a good manner. Specially, the crime occurrence calendar heat map is really attractive. Moreover, we searched several visualizations about crime calendar, there are enough examples that display crime calendar using day of week and hour data. Thus we think display crime calendar using day of week and hour data is meaningful. Next, we tried our best to draw out this kind heat map in our project.

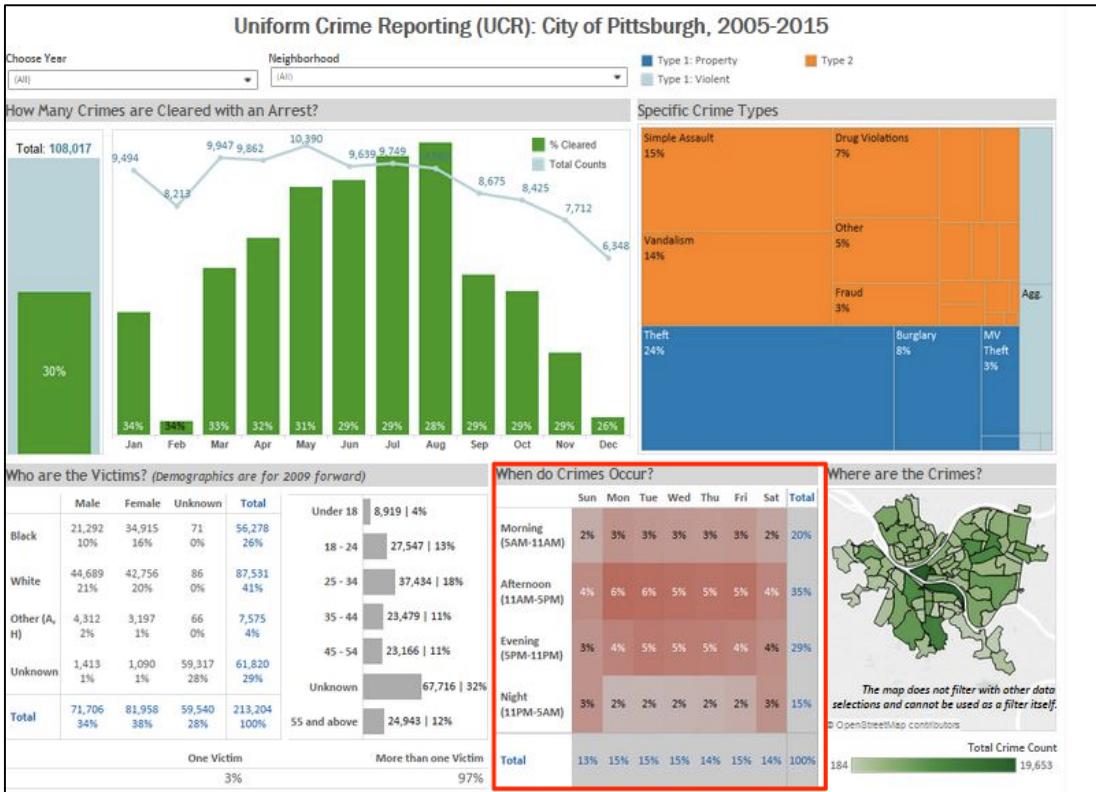


Figure3

For the issue of how to design our website, we also searched some websites of crime topic, like following figure, we think we can also design our website in this style. We think this style is good looking and concise.



Figure4

### 3. Objectives:

Our project involves below objectives below:

- What's the trend of total crime incidents happened in Boston neighborhood from 2012 to 2015?
- Which district in Boston neighborhoods is the most safe and unsafe?
- Which street across Boston and neighborhoods is the most safe?
- What does the crime calendar look like (Month, Day, time) for each street?
- What's the weight for different type of crime? Was it with firearm, unarmed or other? If it was with firearm?
- What's the detailed trend of crime incidents happened in different district, year and month?

From this project, we want to show the data comprehensively. We will visualize the data focused on three different parts:

- First, we will focus on time trend to show the crime frequency.
- Second, we will focus on location related crime frequency.
- Last, we will focus on weapon types used in crime incidents.

In the website, we will give polices several different variables of data that they may want to know.

## **4. Dataset:**

### **4.1 Source:**

<https://data.cityofboston.gov/Public-Safety/Crime-Incident-Reports/7cdf-6fgx>

### **4.2 Data:**

Crime\_Incident\_Reports.csv, 12.3MB, 98,791 records, 11 variables.

heatmap.csv, 37.7MB, 790,609 records, 5 variables.

word\_cloud.csv, 106KB, 4328 records, 3 variables.

### **4.3 Variables:**

For Crime\_Incident\_Reports.csv data:

- a. ID(int): the record number, it is unique for each record.
- b. FROMDATA(int): the time in x/xx/yyyy x:xx:xx AM/PM format when crime happened.
- c. WEAPONTYPE(string): use Firearm, Knife, or Personal Weapon, e.g., foot, hand, etc.
- d. Year(int): from 2012 to 2015, indicates which year the crime happened.
- e. Month(int): from January to December using number 1 to 12.
- f. DAY\_WEEK(string): Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday.
- g. STREETNAME(string): streetname where the crime happened.
- h. ReptDistrName(string): the district name where the crime happened.
- i. lng(float): longitude of the location where the crime happened.
- j. lat(float): latitude of the location where the crime happened.

For heatmap.csv data:

- a. ReptDistrName(string): the district name where the crime happened.
- b. STREETNAME(string): streetname where the crime happened.
- c. DAY\_WEEK(int): 1, 2, 3, 4, 5, 6, 7.
- d. HOUR(int): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24.
- e. Freq(int): frequency for each hour in each day in each street.

For word\_cloud.csv data:

- a. ReptDistrName(string): the district name where the crime happened.
- b. STREETNAME(string): streetname where the crime happened.
- c. Freq(int): frequency for each street in each district.

### **4.4 Clean Up:**

- a. Drop irrelevant variables (not useful for analysis): COMPNOS (Complaint number), NatureCode, X, Y, REPORTINGAREA, DOMESTIC.
- b. For Location, expand it to lat and lng variables and delete the Location variable.
- c. Converting the Year and Month variables from numeric to factor.
- d. Replacing the date time with R-formatted date string, i.e., replace "mm/dd/yy xx:xx" to "yyyy-mm-dd xx:xx:xx".
- e. Deriving the time of the day the crime took place from variable FROMDATE.
- f. Deriving whether the crime took place during the day or at night (7:00pm-7:00am).
- g. Cutting a day into four chunks of interval with 6 hours each.

- h. Reveal the order of DAY\_WEEK, following 'Sunday', 'Monday',..., 'Saturday' sequence.
- i. Delete the data whose lat=0, lng=0, because it will influence the geography visualization; deleting rows whose lat and lng value are NA.
- j. Adjust the letter case confusion.
- k. Converting the REPTDISTRICT into readable Boston district names: i.e., A1='Downtown', A15='Charlestown', etc. And replace REPTDISTRICT to ReptDistrName; deleting rows whose RepDistrName value is NA.
- l. Deriving table from STREETNAME to form street dataset, delete the data whose street.Name is blank.
- m. Remove UCR Part 3 crime records.
- n. Derive frequency for each hour in each day in each street; frequency for each street in each district.

## 5. Exploration, Implementation and Evolution

### 5.1 Part1--Interactive Information Of Streets In Each District:

The first part is focused on streets' information.

We will design a horizontal bar chart (Figure1) shows the crime frequency in Boston and neighborhoods. It would be linked by mouse over one bar to a word cloud figure shown below bar chart(Figure2) shows most frequent streets in each district.

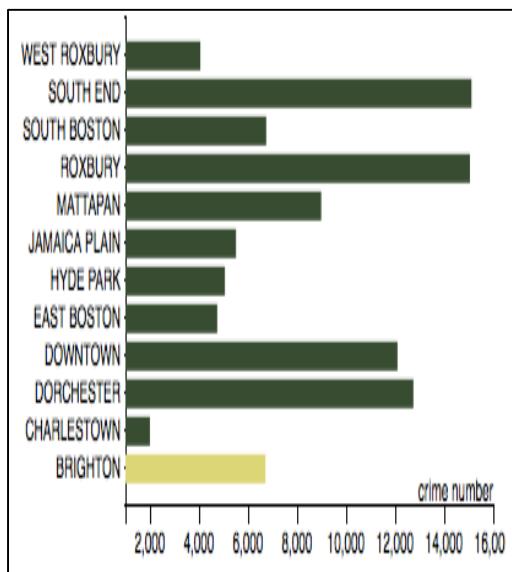


Figure5

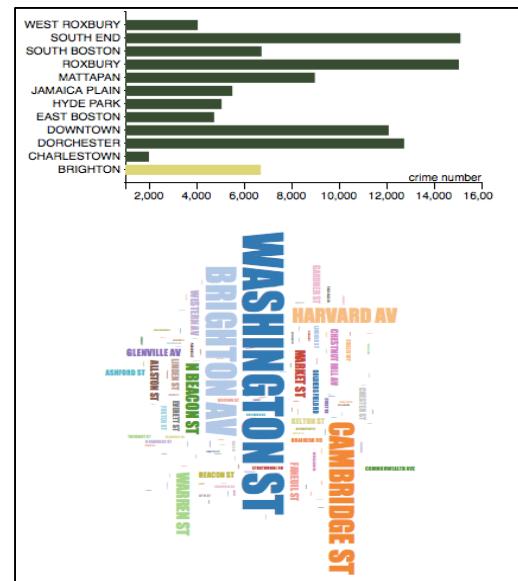


Figure6

Then on the left side of bar chart is a vertical bar chart (figure3) of each month of each street. It would be linked with horizontal bar chart and shown by choosing street names in the dropdown button on top of vertical bar chart.



Figure7

Below the vertical bar chart is a crime calendar heatmap (figure4) with 7 days with 24 hours. It explains the frequency of each day in 24 hours. It also contains the content about weapon type frequency.

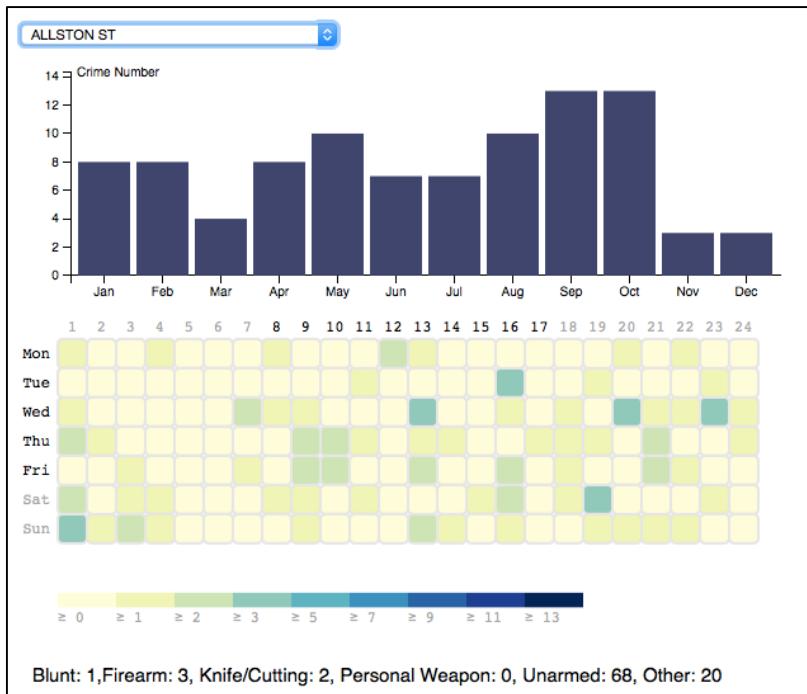


Figure8

Here, we show the overall figures (figure5) below again:



Figure9

#### Summary:

Police can know the detailed information from comprehensive aspects from district, street, month, day and even hour. For example, in Jamaica Plain district, the Columbus Av is the most dangerous street; and in Columbus Av, the most dangerous month is December, and on Friday, the most dangerous hour is 22:00. In this street, it happened with 6 firearm crimes , 6 knife/cutting crimes. So police can have a rough sense of Columbus Av street in Jamaica Plain.

#### 5.2 Part2--Suggestion To Police On Different Weapon Types:

We also give some suggestion on each weapon type. We think it is necessary to let people and also help police how to give suggestion on each weapon type.

Gentle Suggestions	
Weapon Type	Tips
Blunt	<p>The choices are simple... Run away, or close the distance and strike while defending.</p> <p>Staying in range of the stick, but out of range to counter attack, is potentially the biggest mistake you could make in this situation. Remember, the stick will be moving fastest at the point furthest from the attackers arm, so getting inside that point on the stick is crucial.</p>
Firearm	<p>Don't touch! Leave the area.</p> <p>Indoors, the overpressure from firing the gun has nowhere to go and is reflected back at you from the walls. It can be conducted to your inner ear through the mastoid bone and do damage even though you don't feel anything. You should probably use both muffs and plugs simultaneously and always get the highest rated ones you can find.</p>
Knife	<p>If you can see the knife in hand and the attacker is some distance away – widen that distance, run away!</p> <p>If there's no where to run or you can't leave the situation for some reason, then get something solid between you and the attacker (chair, baseball bat, rucksack, anything – just enough to distract him, so you can go to work on him).</p> <p>Take careful note where his hands are and what he is doing with them.</p> <p>If his hands are empty and you get in to a scuffle, he may reach in to his pocket and take a blade out and get use it – So, watch those hands – again if you have any training – trap the arms/hands while striking him.</p>
Personal Weapon (hand foot etc.)	<p>If a robber asks for your wallet and/or purse, DO NOT HAND IT TO HIM.</p> <p>Toss it away from you....chances are that he is more interested in your wallet and/or purse than you, and he will go for the wallet/purse. RUN LIKE MAD IN THE OTHER DIRECTION!</p> <p>If you are ever thrown into the trunk of a car, kick out the back tail lights and stick your arm out the hole and start waving like crazy. The driver won't see you, but everybody else will. This has saved lives.</p>
Unarmed	Still Watch out, please! :-)

Figure10

### 5.3 Part3--Statistical Information On Crime Incidents:

This part will provide more detailed of crimes data.

At first, a histogram of crimes from year 2013 to 2015 appears first, then once you click one of them, a table of districts with crime number in ascending order will appear on right hand side.

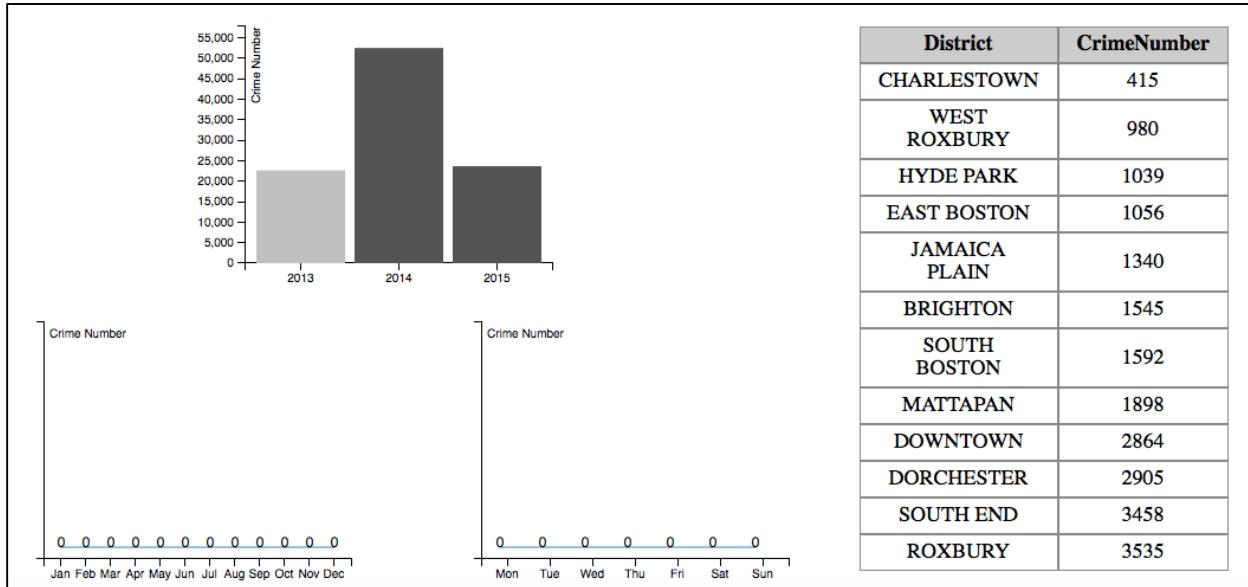


Figure11

Second, once you click each district name in the table, the district name's color will change to red; at the same time, two line charts will appear under histogram graph after clicking the district name in that table. Left hand line chart is month trend, and right hand line chart is day trend.

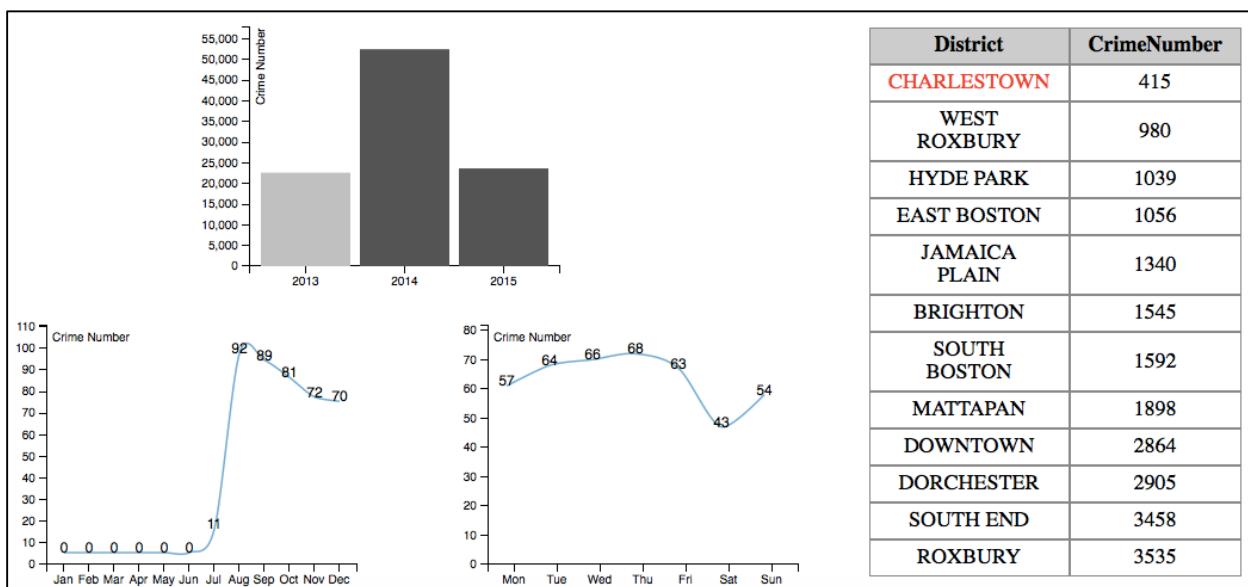


Figure12

Police can get the sense on each district of what time is with high crime frequency.

And they can also know general trend of crimes in Boston and neighborhoods. In our analysis, police can easily and directly know there are much more crimes happened in 2014 than 2013 and 2015. And in 2013, Charlestown district is the most safe area, plus, most crimes happened in the last 5 months.

## 6. Design Evolution:

Initially, our group came up with four patterns of visualization to show our project, each pattern involves three parts. Here, we would introduce them briefly as following:

### 6.1 Pattern1:

Figure9: it shows the relation between crime and location, we want to design a map (figure 1) by marking districts as dots on map, by clicking the dot on map, there will be a table (figure 2) on the right shows the detailed information of each crime.

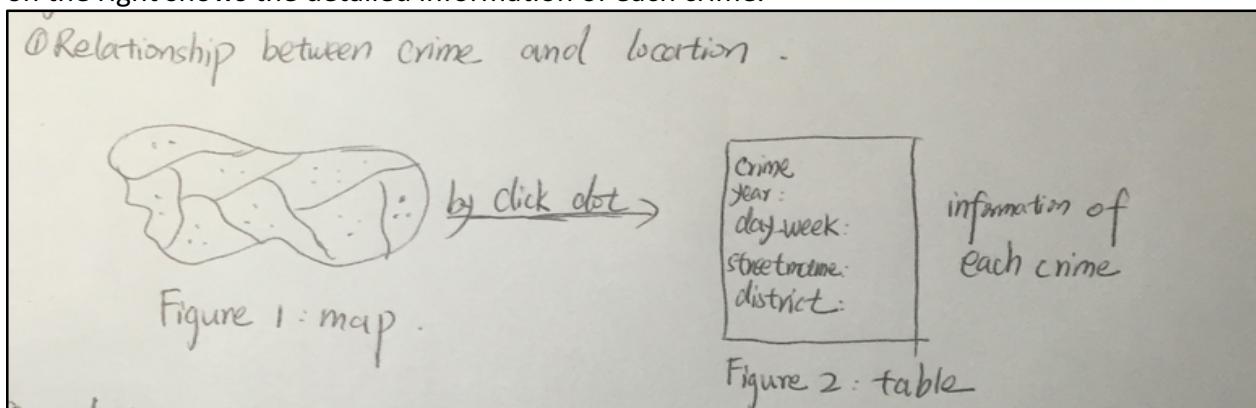


Figure13

Figure10: it is to show the different crime rate in different districts. There will be a map showed the districts (figure 3) with gradient color, a color legend is on the right.

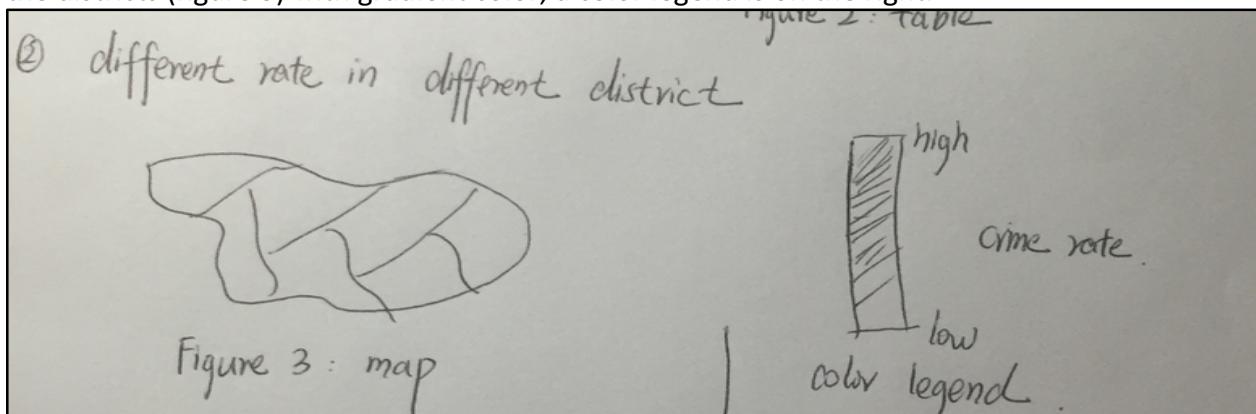


Figure14

Figure11:it is to show the crime numbers in each district. By clicking the map (figure 3), a vertical bar chart (figure 4) will show up to show crime numbers in different years. And by clicking the bar, a pie chart (figure 5) will show up to show the crime numbers of different

month for each year.

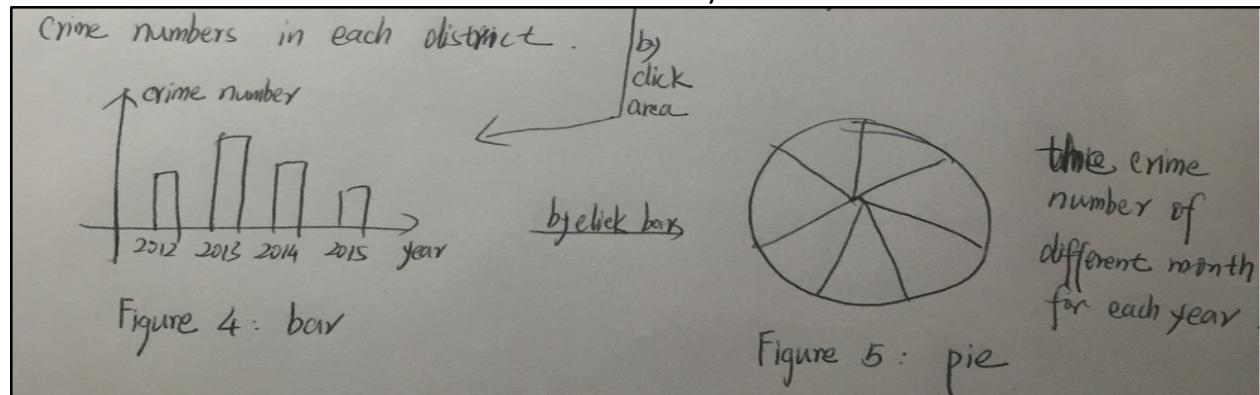


Figure15

## 6.2 Pattern2:

In this pattern, we aim at design a website contain contents for different website lookers. One for government, one for statisticians, and the other one for newsletters.

Figure12: it is for government, we want to do a bubble line chart (figure 1) to show the total crime numbers in different year, and by clicking the bubble, a pie chart (figure 2) will show up to show the crime number for each district of each year.

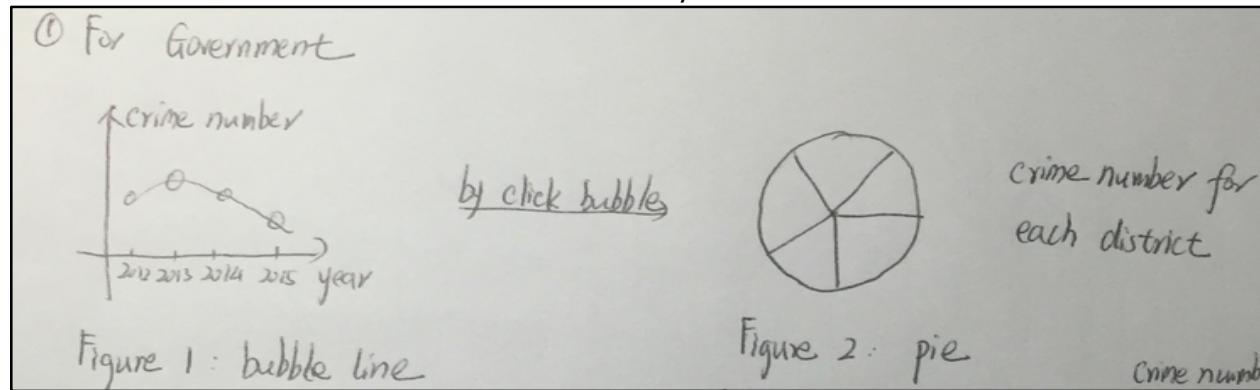


Figure16

Figure13: it is for statistical researcher, we do a district map (figure 3), by clicking the district, a table (figure 4) contains the name of district and total crime numbers of district will show. By clicking the district name, a horizontal bar chart (figure 5) shows the total crime numbers for each month and another horizontal bar chart (figure 6) shows the total crime numbers for each day will show up simultaneously.

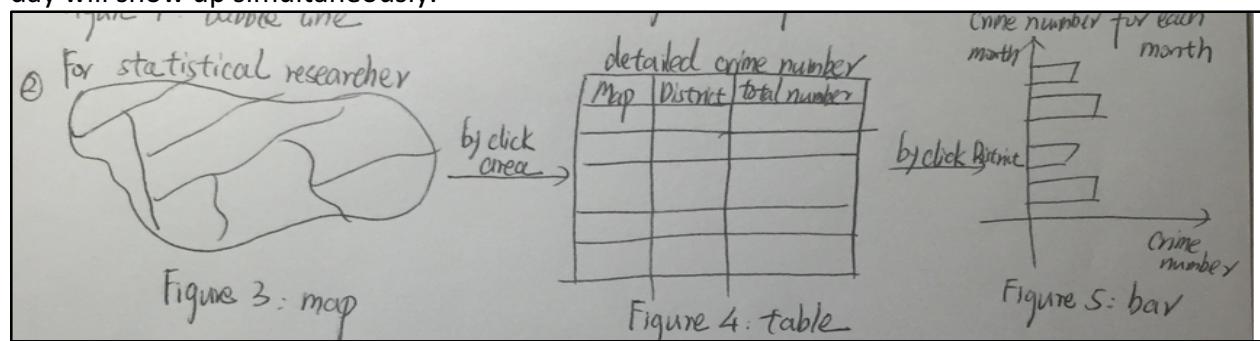


Figure17

Figure14: it is for newsletters, a pie chart (figure 7) will show up with the information of whether the crime is with weapon or not. By clicking the firearm area, a vertical bar chart will show up to show the detailed shooting number when with firearm for each district.

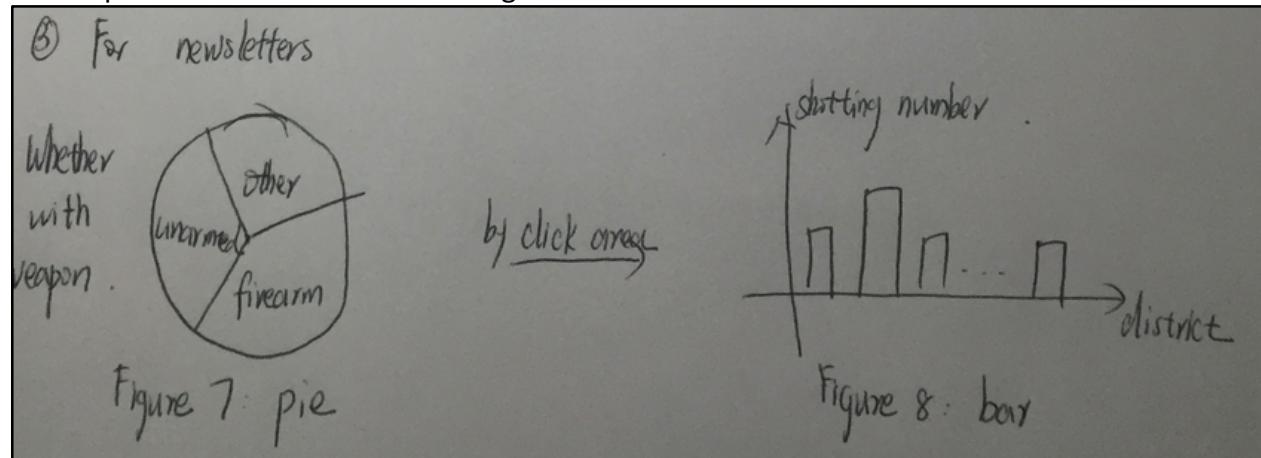


Figure18

### 6.3 Pattern3:

Figure15: it is "District Detailed Profile", for each district, draw STREETNAME, Month, DAY\_WEEK on parallel coordinate. When click on one line in parallel coordinate, the related dashboard will display on left side, which contains WEAPONTYPE, Shooting, timeDec. From this part, we want to show which month and day relate closely with one specific street. And for specific street, what's the weapontype, whether shoot, when to happen.

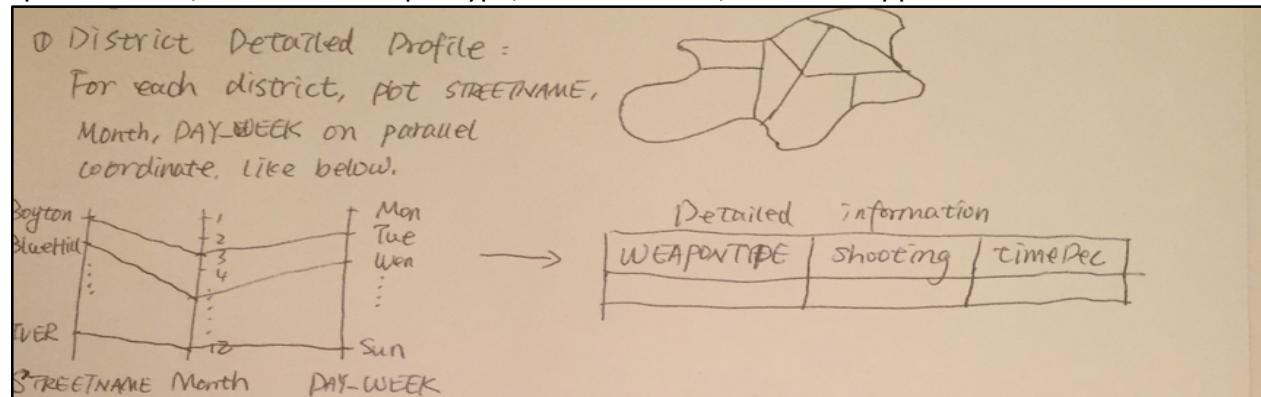


Figure19

Figure16: it is the suggestion based on the first part. A dashboard contains three columns: ReptDistrName, Month, DAY\_WEEK. The first column consists of the top3 safest district; and for each district, which months are the safest, which DAY\_WEEK are the safest.

② Suggestion =  
we will give advice for which district, month, DAY-WEEK would be taken care.

Suggested District	Month	DAY-WEEK
District1 (top 3)	--	--
District2	--	--
District3	--	--

Figure20

Figure17: it shows the crime trend. This trend visualized by bar chart. For interactively displaying, we design that by clicking each bar, two wordcloud graphs will appear on left side. These wordcloud graphs are based on the STREETNAME frequency and WEAPONTYPE respectively.

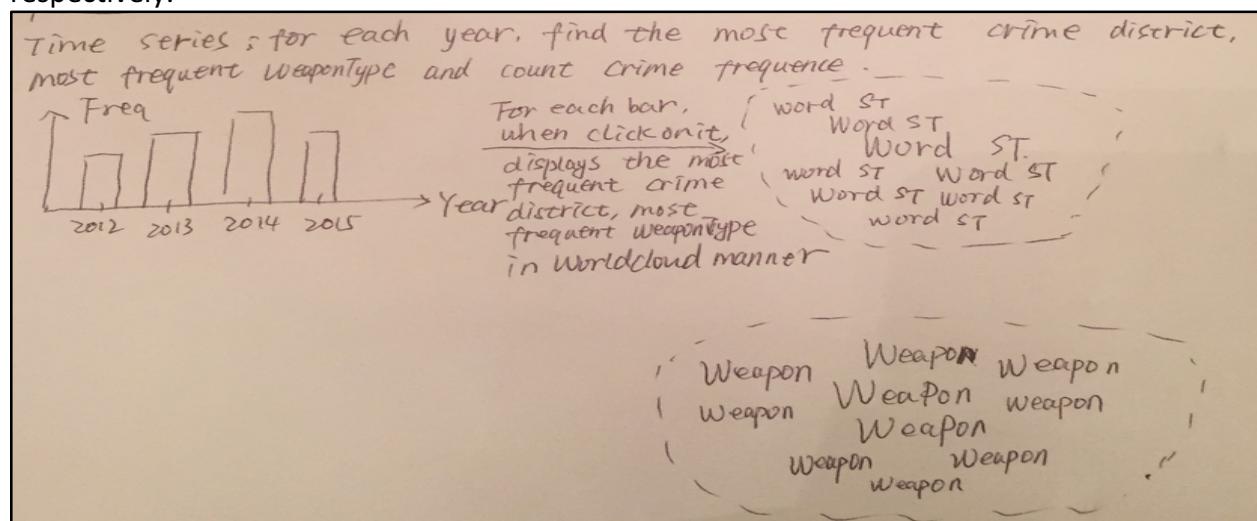
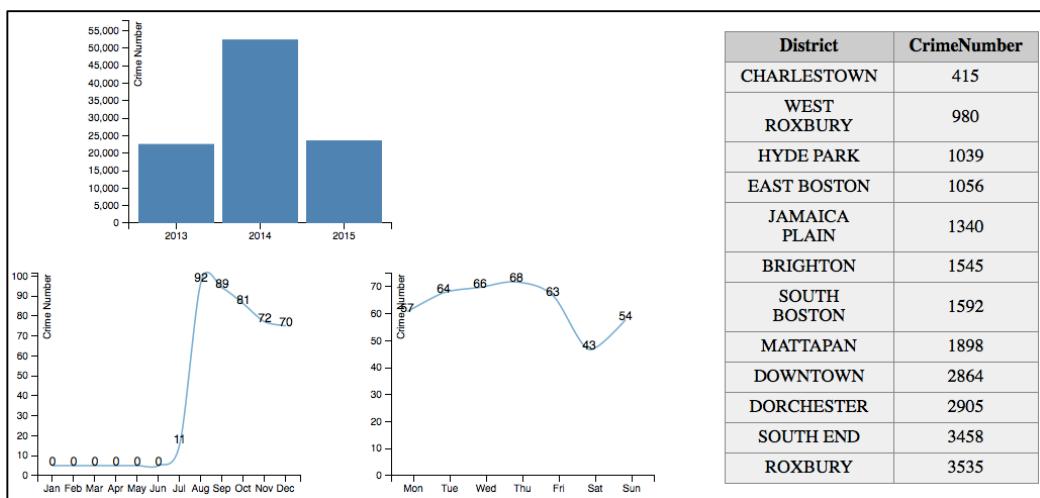


Figure21

#### 6.4 Pattern4:



Figur22

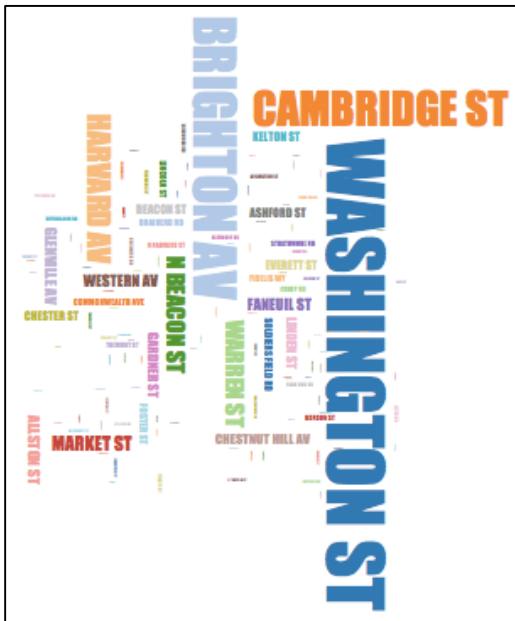


Figure23

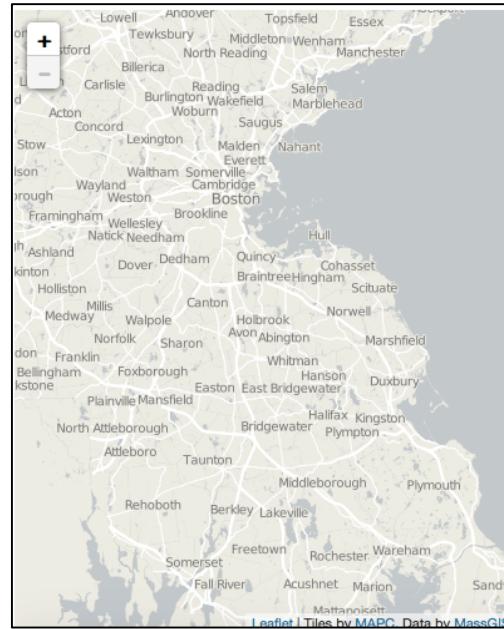


Figure24

**Authors:** Yu Tian, Naihui Wang, Yang Liu

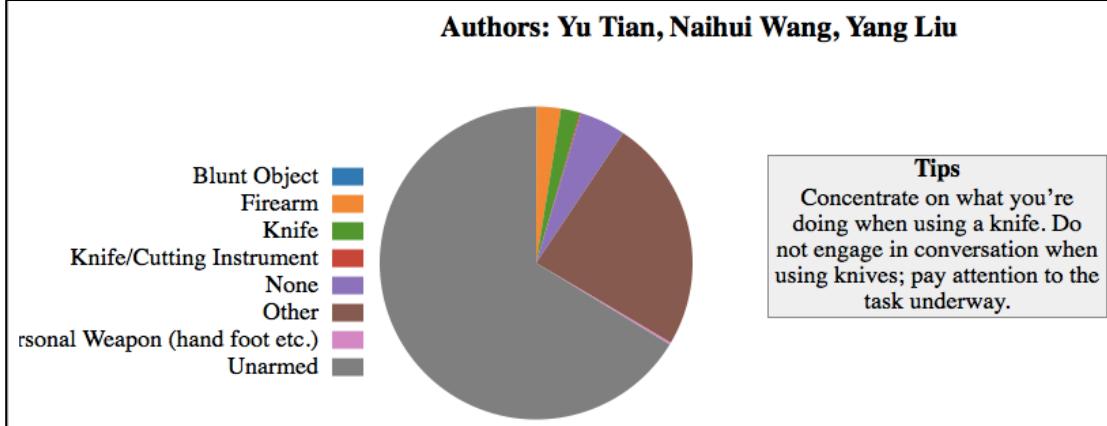


Figure25

## 7. Evaluation:

### 7.1 Exploration of Data:

- ❖ Time period is from 07/29/2013 to 08/10/2015.
- ❖ The top5 unsafe districts are: SOUTH END, ROXBURY, DORCHESTER, DOWNTOWN, MATTAPAN.
- ❖ For each street in each district, the crime frequency in each month is in much difference.
- ❖ Year 2014 has the most frequent crime incident.
- ❖ Charlestown and West Roxbury are always the most safe districts from year 2013 to year 2015.
- ❖ In both year 2013 and 2014, the frequency of crime in each month does not have much difference.
- ❖ Most crime incidents are unarmed.

## **7.2 Future Work:**

Come up with better visualization type to explain our data like:

- ❖ Make map visualization.
- ❖ Introduce GPS to locate where the user is, and then calculate the distance to each destination, finally, calculate the crime frequency in a specific distance, i.e., 0.3 mile to the destination where user want to go.

## **8. Reference:**

Dataset Source:

<https://data.cityofboston.gov/Public-Safety/Crime-Incident-Reports/7cdf-6fgx>

Design Inspired By:

<http://www.city-data.com>

<http://www.crimeinchicago.org/>

<http://www.dataplusscience.com/insights.html>

<http://opendatabits.com/visualizing-the-relationships-between-chicago-homicides-and-hot-weather/>

<http://public.tableau.com/profile/jeffs8297-1/vizhome/ACincinnatiCrimeVisualization/StartPage>

<https://www.wprdc.org/showcase/city-of-pittsburgh-crime-visualizations/>