

Exploration Of Crime In Boston

Tian Yu(ytian2@wpi.edu, github: ytian2)
Naihui Wang(nwang@wpi.edu, github:Naihui)
Yang Liu(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, 15th, 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. Crime incidents are just around us, in our daily lives. The world is never in peace, even in Boston neighborhood, crime incidents are everywhere, affects our daily life and safety.

Under this circumstance, we decide to do this project about crime incident in Boston neighborhood. From this project, we are trying to let people 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 let everyone be cautious and know he or she should always be prepared for the crime. Even if there is only one person can be more alert about crime from our visualizations, we can say we did a good job, this visualization is useful.

2. Objectives:

- a. What's the trend of total crime incidents happened in Boston neighborhood from 2012 to 2015?
- b. Which district in Boston neighborhood is the most safe and unsafe?
- c. What's the weight for different type of crime? Was it with firearm, unarmed or other? If it was with firearm, did shooting happen?
- d. What's the detailed number of crime incidents happened in different district, year and month?
- e. Which year, month did crime happen most?

From this project, we want to show the data deeper and deeper. 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.

From the website, we will give people several different variables of data that they may care about.

3. Dataset:

3.1 Source:

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

3.2 Data:

Crime_Incident_Reports.csv, 12.3MB, 98,791 records, 11 variables.

3.3 Variables:

- a. ID(int): the record number, it is unique for each record.
- b. WEAPONTYPE(string): use Firearm, Knife, or Personal Weapon, e.g., foot, hand, etc.
- c. Shooting(Boolean): Boolean variable, shoot or not.
- 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. timeDec(string): four time intervals for one day, i.e., (0,6], (6,12], (12,18], (18, 24].
- i. ReptDistrName(string): the district name where the crime happened.
- j. lng(float): longitude of the location where the crime happened.
- k. lat(float): latitude of the location where the crime happened.

3.4 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.

4. Exploration, Implementation and Evolution

4.1 Part1--Interactive Bar to Table to Line:

The first part is focused on time trend of crime numbers.

We will design a bar chart (Figure1) shows the crime frequency from 2012 to 2015 in Boston neighborhood. It may link by mouse over one bar to a table (Figure2) shows the crime frequency of different district each year. Then the table will link by mouse over one row to two line charts (Figure3), one shows the crime frequency for every month in different district. Another shows the crime frequency for every day of week in different district.

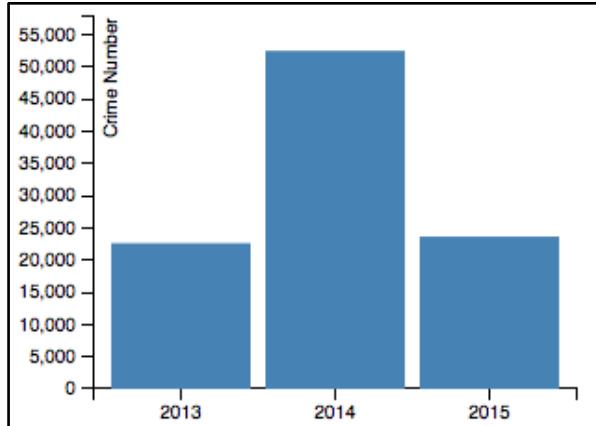


Figure1

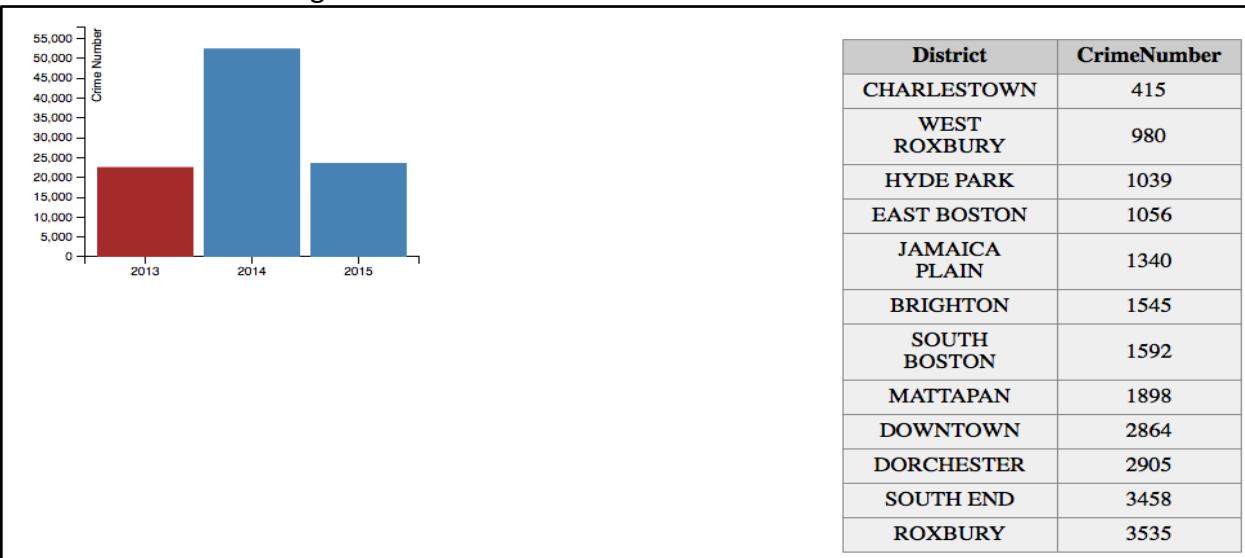


Figure2

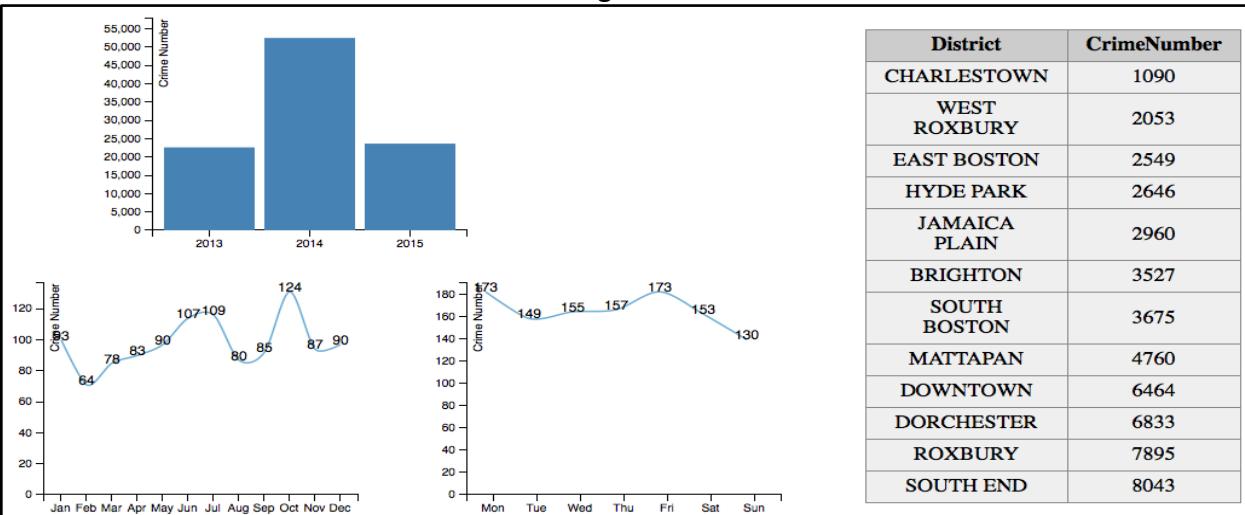


Figure3

4.2 Part2--Interactive Word Could to Map View to Parallel Coordinate:

The original idea of the second part involves three visualizations: word cloud, map, parallel coordinate. On left side, we draw a word cloud (Figure4) of ReptDistrName to know which districts are most not safest: the large the size of word is, the worse the safety is. Then we will design a district map (Figure5) to further exploration in detail by mouse over one district to parallel coordinate (Figure6) involves STREETNAME, Month, DAY_WEEK, and related Frequency.

By drawing parallel coordinate, we can easily get the story that in specific district, the frequency of specific streets in specific month on specific day of week. Then people can easily cluster those streets by month and day of week.

However, we want to change our idea on parallel coordinate because our Geojson data can only show each point in map instead, not the district. So we do not draw word cloud for district, instead we draw word cloud of streetname. Consequently, the parallel coordinate has no meaning here. We are thinking of other types.



Figure4



Figure5

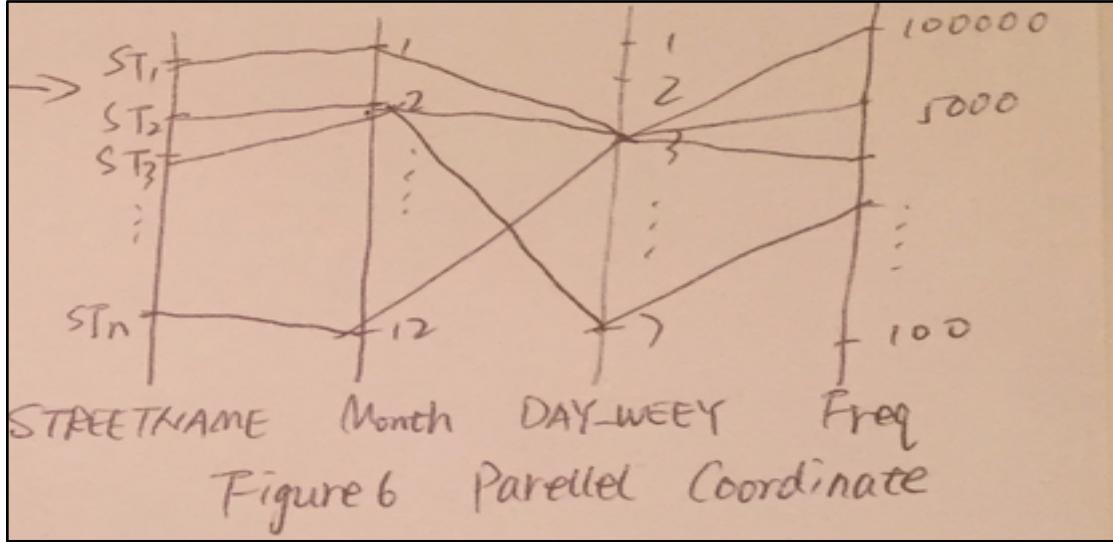


Figure6

4.3 Part3--Pie to Table:

The third part is focused on weapon type used in every crime incident.

We will design a pie chart (Figure7) shows the frequency of different weapons used in all crime incidents. This pie chart will link to a table (Figure8) by mouse over part of pie chart give some tips for self-protection when faced with different weapons in crime.

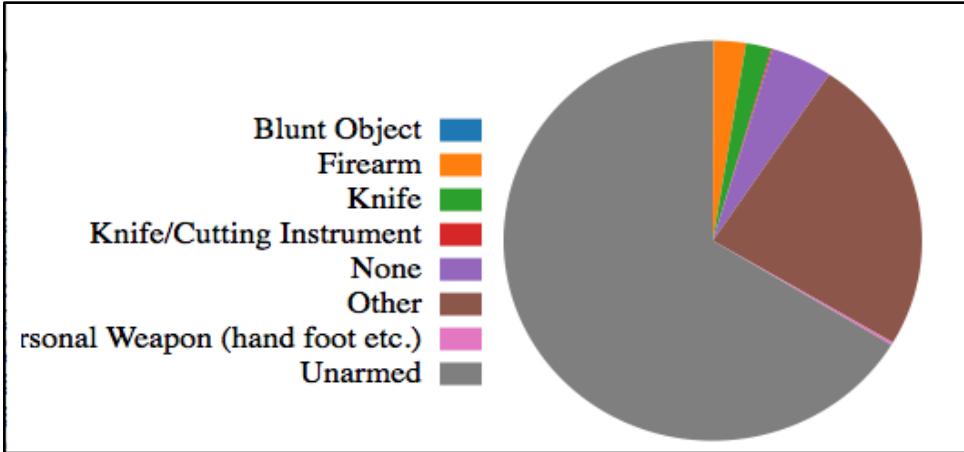


Figure7

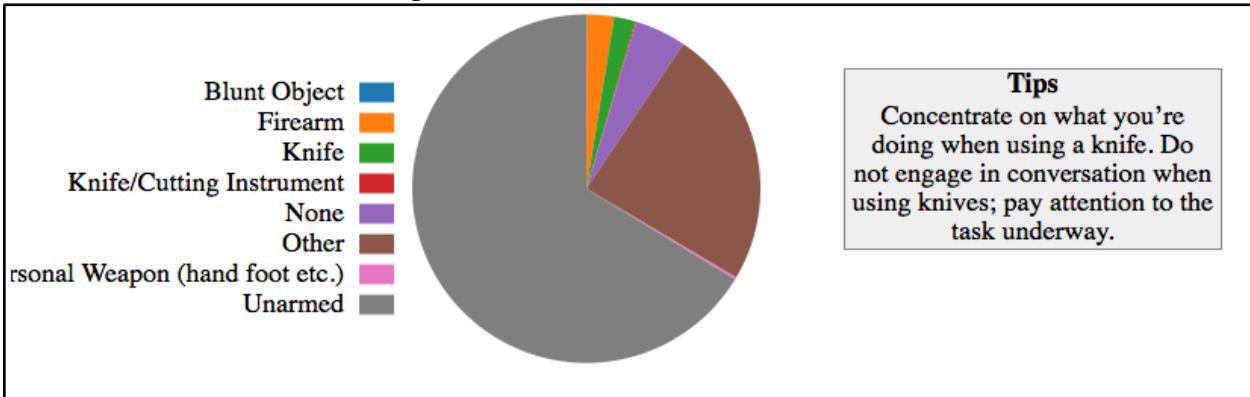


Figure8

5. Design Evolution:

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

5.1Pattern1:

Figure9: it shows the relation between crime and location, we want to design a map (figure 1), the crimes are marked 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.

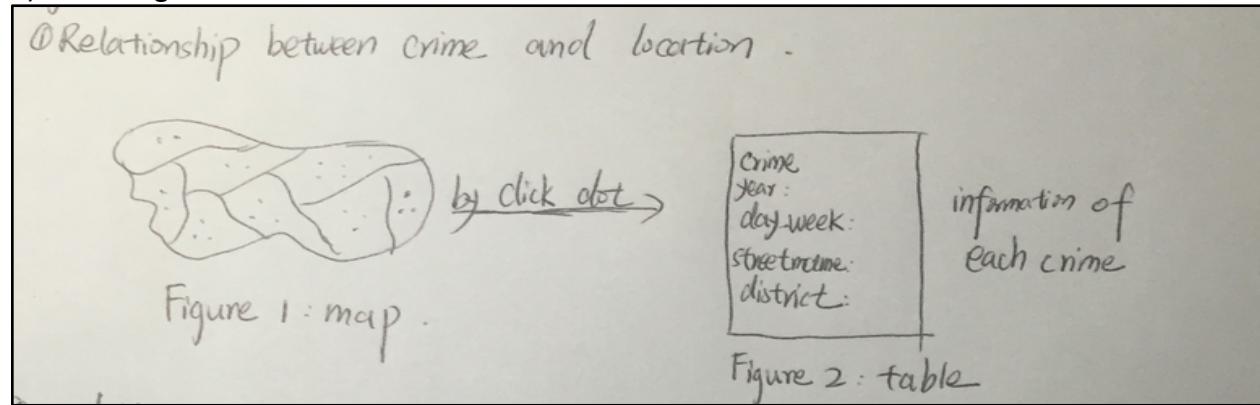


Figure9

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.

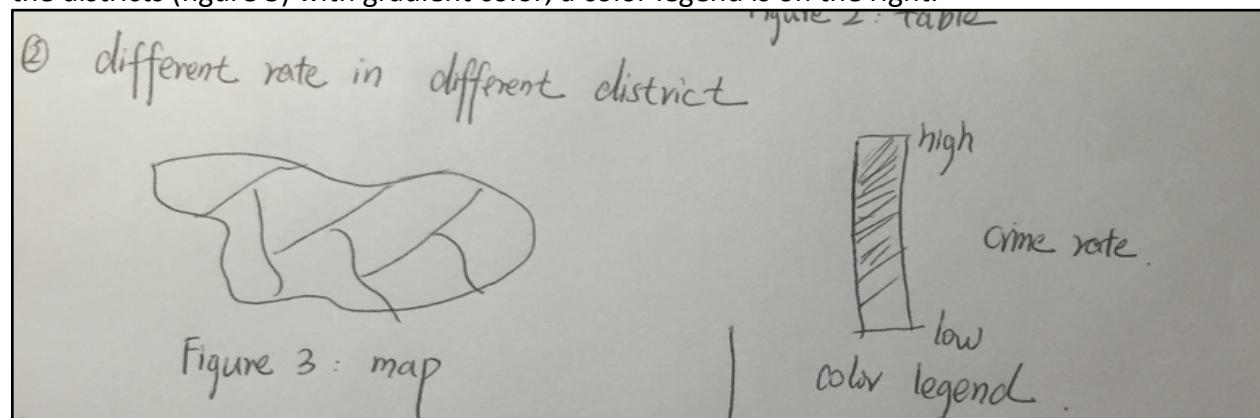


Figure10

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.

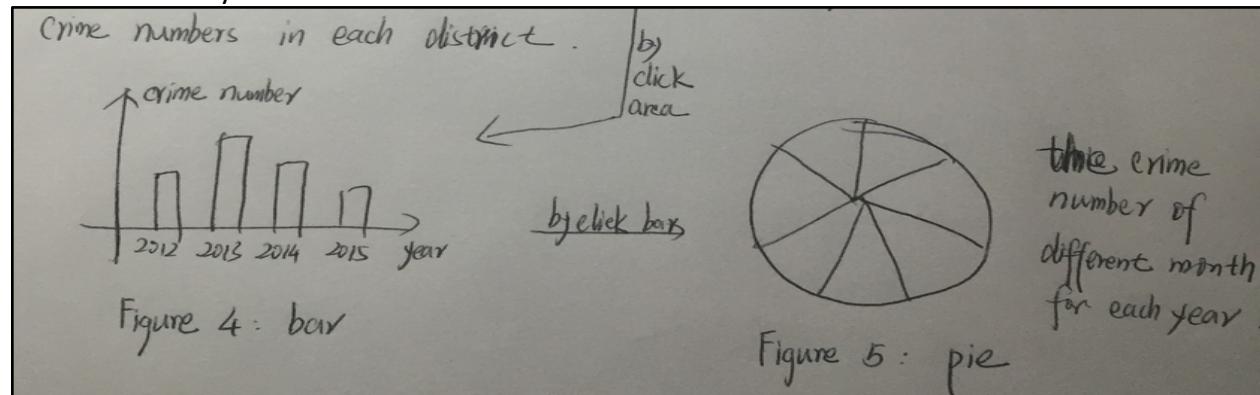


Figure11

5.2 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.

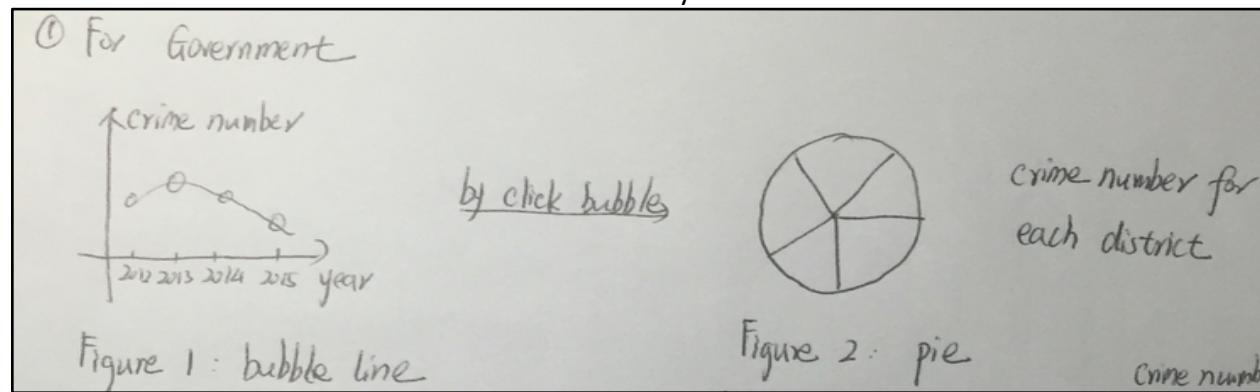


Figure12

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.

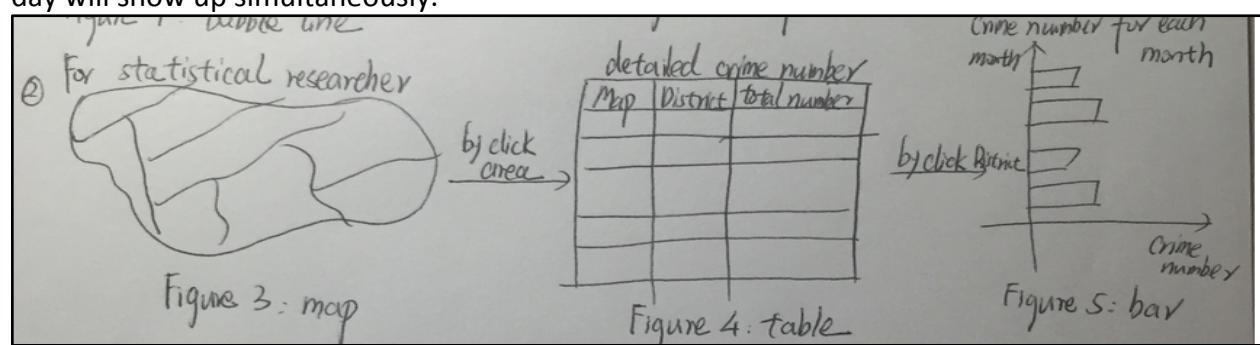


Figure13

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.

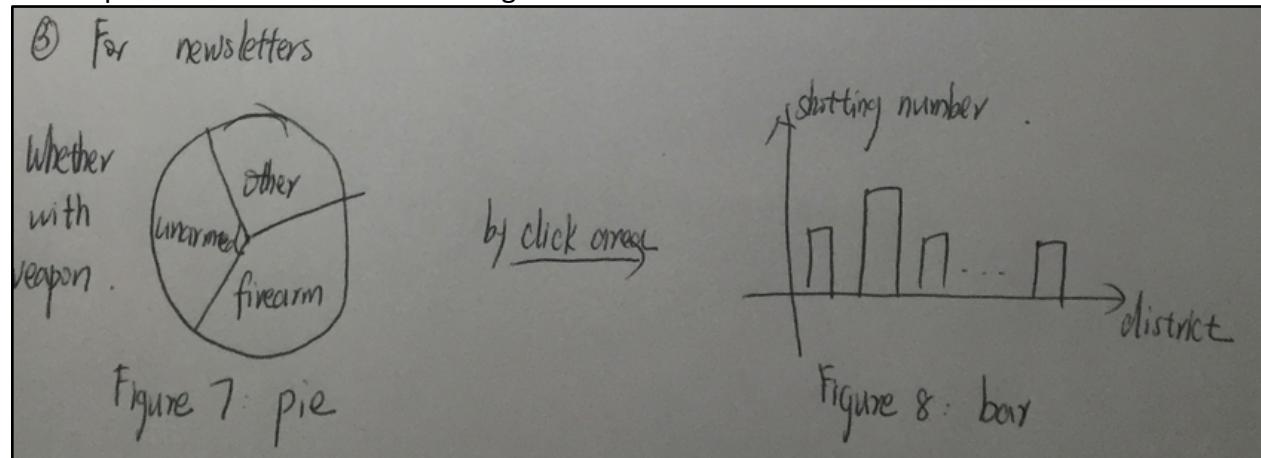


Figure14

5.3 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.

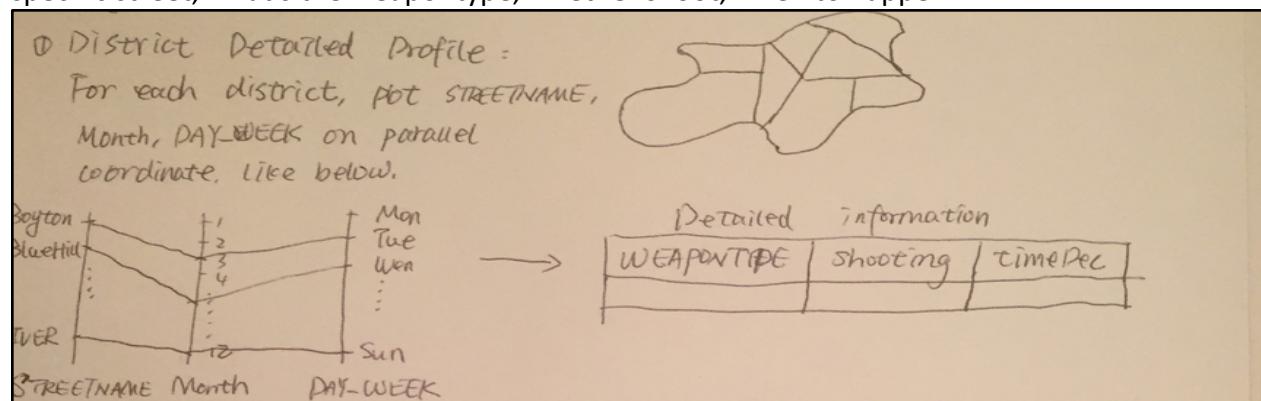


Figure15

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 (top3) | --- | -- |
| District2 | -- | -- |
| District3 | -- | -- |

Figure16

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.

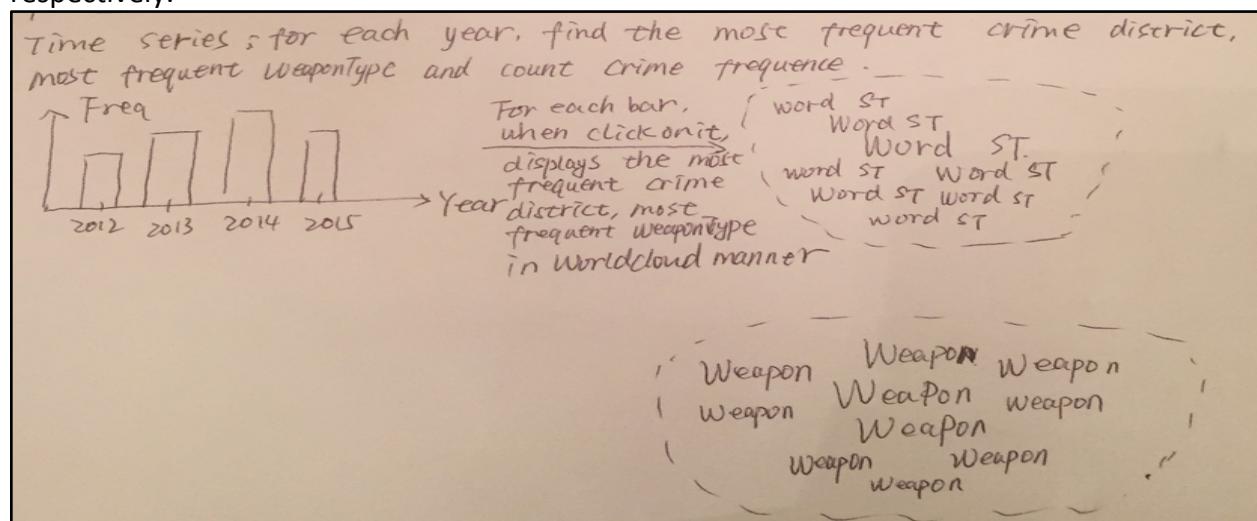


Figure17

6. Evaluation:

6.1 Exploration of Data:

- ❖ The top5 weapontypes are: unarmed, other, none, firearm, and knife.
- ❖ Year 2014 has the most frequent crime incident.
- ❖ Charlestown is always the most unsafe district from year 2013 to year 2015.
- ❖ The top5 unsafe districts are: SOUTH END, ROXBURY, DORCHESTER, DOWNTOWN, MATTAPAN.
- ❖ The frequency of crime in each day of week does not have much difference.
- ❖ In both year 2013 and 2014, the frequency of crime in each month does not have much difference, while in year 2015 before August, the frequency of crime is much larger than after August.

6.2 Future Work:

- ❖ Modify the visualization in part 2. Come up with better visualization type to explain our data.
- ❖ We will aggregate the three parts into one and show in one webpage.

7. Reference:

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

Design Inspired By: <http://www.city-data.com>