The Crime Landscape of Salt Lake City

Basic Information

Members:

- Archit Rathore – u1144416 (archit@cs.utah.edu)
- Yash Gangrade – u1143811 (yashgangrade09@gmail.com)
- Rebeka Mukherjee – u1141112 (rebeka.mukherjee@utah.edu)

Github Repository: https://github.com/yashgangrade09/dataviscourse-pr-crime-landscape.git
Youtube Video Link: https://www.youtube.com/watch?v=a9RgT5zZu4g&feature=youtu.be
Project Website: https://architrathore.github.io/slc_crime_landscape/

Contents

MOTIVATION

The general consensus between us team members was to work with data that pertained to some societal factors. Another thing that we wanted out of our project was to be able to derive insights on a much finer geospatial resolution (state/city). Finally, we also wanted our visualization to be relatable and not dealing with data of a technical nature. This led us to explore the Utah Open Data catalog (https://opendata.utah.gov/) and we finally zeroed in on the police cases dataset that lists all reported crimes in the Salt Lake county.

This dataset checks off all our boxes, namely:

- Pertains to societal factors
- Is localized
- Is relatable and may be of equal interest to both a layman and a specialist
- Has the potential to inform policy decisions

DATA COLLECTION AND PROCESSING

As we discussed before, we are using the Police cases dataset from 2008-2016 from Utah Open Data catalog. Each record in the data corresponds to an instance of reported crime at one of the police stations in Salt Lake City. It contains information about the type, time and date of occurrence and reporting, and location of the crime. There are about 50K+ records per year. The data is open access.

Each year dataset had the following information: Case Number, UCR Code or NCIC Code, Description, IBR Code, Occurred Date, Reported Date, Day of Week, Location, City, City Council, Police Zone, Police Grid, X-Coordinate, Y-Coordinate.

The two main problems with these datasets are:

1. The schema is not the same in each year dataset. For example, some years display the UCR Code for each record while some display the NCIC Code. The column names are also different for every year.
2. Some crime types are not recorded over all the years.
3. The size of each dataset is around 10 MB and trying to load all the data will make the visualization slow.

In order to overcome these problems, we took the following measures:

1. We decided on a schema that would be uniform in data across all the years.
2. We removed all rows in which data from any of the columns in the schema are missing.

3. We considered only a subset of the crimes to visualize which are more significant in the dataset and which are present over all the years from 2008 to 2016.
4. We also grouped certain crimes which were like each other.

We processed each year dataset and create a CSV file named '[year]_processed_nowhitespace.csv' with the following columns:

DESCRIPTION: It is a description of the crime in a particular record. The value is one of the following categories: Assault, Burglary/ Larceny/ Robbery, Damaged Property, Drugs, Homicide, Kidnap, Traffic, Weapons.

The original data had a column for Occurred Date and Reported Date. In most cases the Occurred Date and Reported Date are the same. We considered Occurred Date for finding the valued of **DAY**, **MONTH**, **YEAR** and **TOD** columns for all of the years except 2011, where we considered the Reported Date since Occurred Date was not recorded.

DAY: It is the day of the month that the crime in the record was committed.

MONTH: It is the month of the year that the crime in the record was committed.

YEAR: It is the year that the crime in the record was committed.

TOD: It is the time of day that the crime in the record was committed. The 24 hours in a day are divided into 8 parts: 12 am to 3 am, 3 am to 6 am, 6 am to 9 am, 9 am to 12 pm, 12 pm to 3 pm, 3 pm to 6 pm, 6 pm to 9 pm, and 9 pm to 12 am. The value of **TOD** is between 1 and 8 according to which octant it belongs to.

DOW: It is the day of the week that the crime in the record was committed. This data was directly taken from the original dataset. It is a value between 1 and 7 depending on the day of the week where 1 means the crime was committed on a Sunday.

ADDRESS: It is the street address of the location where the crime in the record was committed. This was also directly taken from the original dataset with very little processing like removing of whitespaces.

In order to populate the Map View, we need the **LATITUDE** and **LONGITUDE**. The X-coordinate and Y-coordinate in the original dataset are State Plane Coordinates and we could not figure out a way to convert them into the corresponding latitude and longitude. Hence, we had to use the street address to find the latitude and longitude. We aggregated all the year datasets and got a list of the unique addresses. Then we tried to use a python package called geopy to find the latitude and longitude for each address. When this method could not find the latitude and longitude of over 3000 addresses out of 9224 addresses, we turned to Google Maps to obtain the latitude and longitudes.

LATITUDE: The latitude of the location where the crime in the record was committed.

LONGITUDE: The longitude of the location where the crime in the record was committed.

These steps reduced the size of the data files drastically from around 10 MB to 2 MB, but we faced trouble again while implementing the statistics view. The main problem was that we had to process the data again to be able to create Line Charts for the Yearly Statistics, Monthly Statistics, Weekly Statistics, and Hourly Statistics from it. If we did the processing in JavaScript, it was taking a lot of time to process and load the data. Hence, we decide to create separate json files called 'all_years.json', 'all_months.json', 'all_weeks.json' and 'all_hours.json'. 'all_years.json' contains the aggregated data over the years for each crime type. 'all_months.json', 'all_weeks.json' and 'all_hours.json' contain the aggregated data over the years for each month of year, day of week or time of day for each crime type.

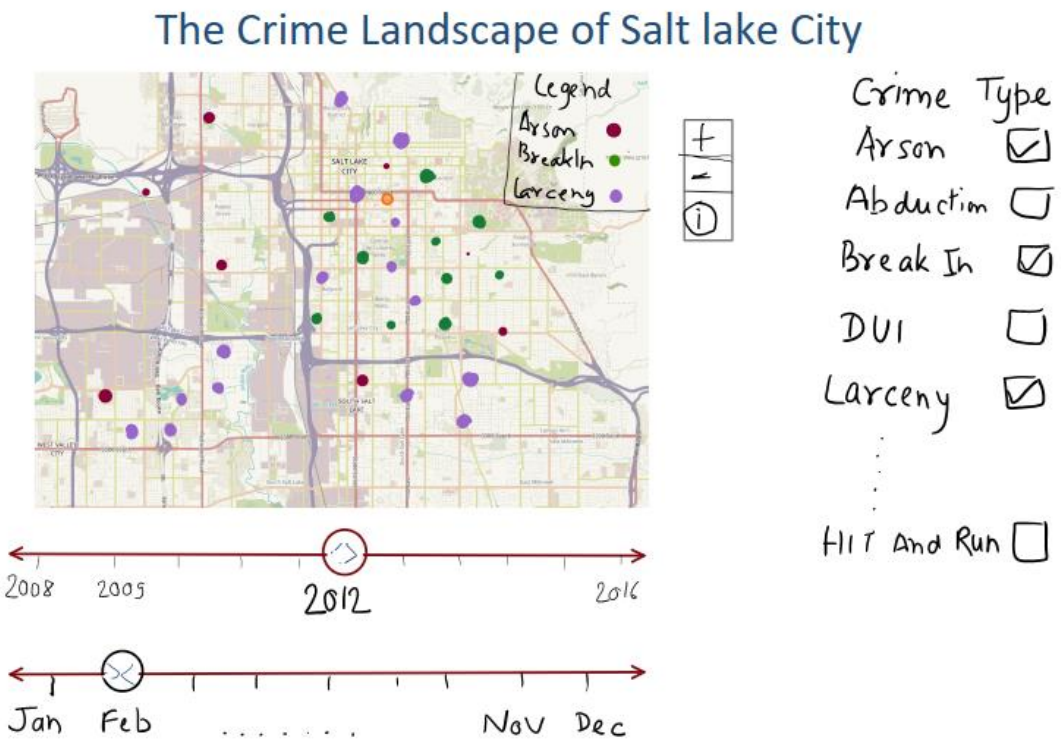
DESIGN EVOLUTION AND IMPLEMENTATION

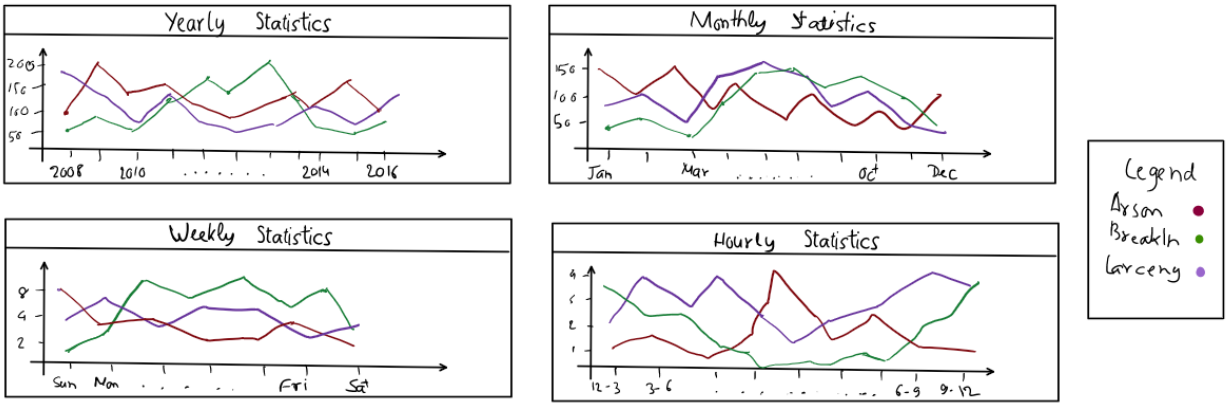
PROPOSAL DESIGN:

In our main proposal design, we had three different views namely Map View, Selection View and Statistics View to show the yearly, monthly, weekly and hourly statistics. Details of the design are as follows:

- The map view shows the location points of crimes selected in selection view and time selected by the time slider. The mark used is points and channels are color (type of crime) and position (location of crime).
- A time slider enables the user to see the trend of crimes over the years. On clicking the year slider, it expands to show the months for the selected year. This enables the user to visualize the crime trends on a finer scale. It collapses to show only the years when clicked again.
- A selection view enables users to select the crimes they wish to visualize through checkboxes.
- The map view has a semantic zooming feature with zip code level aggregation at the topmost level and blocks at the lowest level.
- The statistics view shows yearly, monthly, weekly and hourly statistics of aggregated data about selected crimes as line charts.
- All the views will be linked to each other. For example, clicking on crime selection view will update the map view as well as the statistics view.

Our proposal design looked something like this:





Apart from the main design, we also submitted three prototypes of alternative designs which might be useful. We are using one of the prototype designs and creating a pie chart which will demonstrate the distribution of crimes over a year.

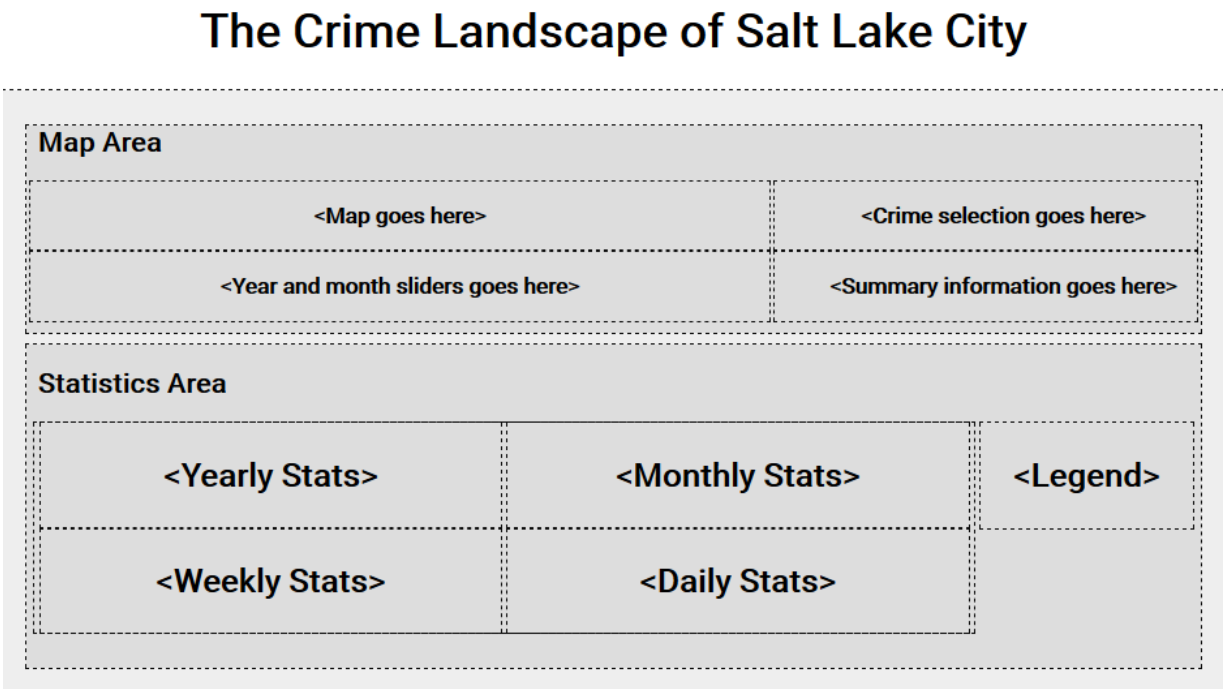
CHANGES AND IMPROVEMENTS IN DESIGN OVER THE COURSE OF PROJECT

We made quite a few changes in the design with respect to all the views.

1) Page Structure

PAGE STRUCTURE

We first created the basic structure of the directory and created the views structure of the visualization. First, we created a container which holds all the views and structure in several nested divs. Each div represents a view like map-view, summary view etc. and we are assigning classes to each of the divs. This assists us in arranging the divs according to the structure required. Also, we can work on a view separately without disturbing the entire structure. Below is the initial starting structure of the visualization.



YEAR SLIDER

The visualization needs a year slider to update the map when the year is changed. We directly used the year slider structure from the Gap Minder HW here. Similar slider will be attached for the month data as well. An image of year slider is attached below.

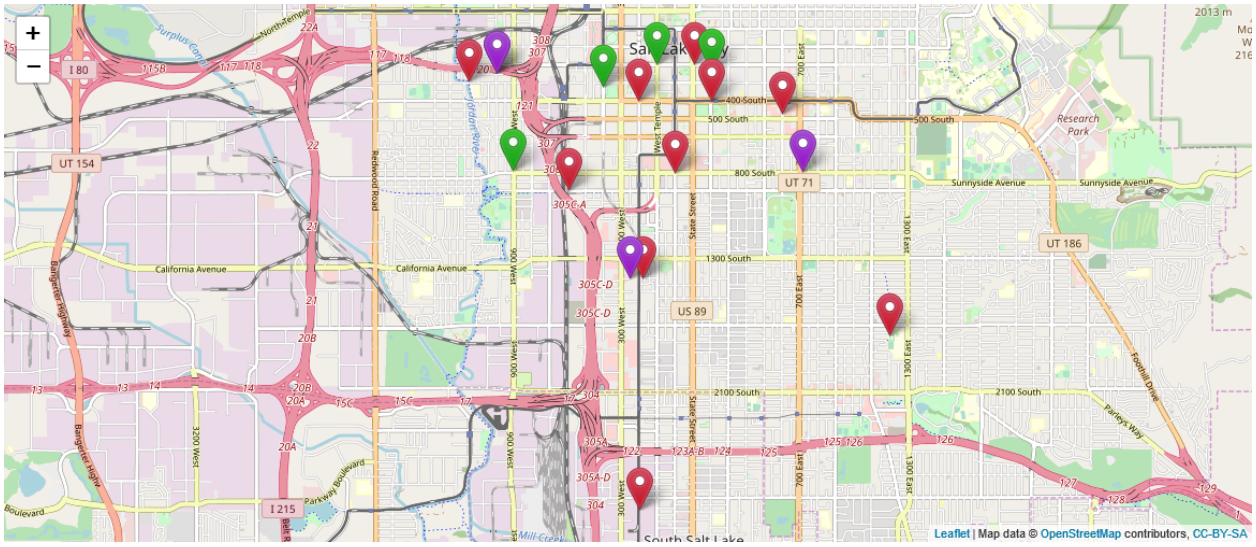
2) Map View

For each year, we had nearly 60000 rows in the data i.e. ~60K crimes per year.

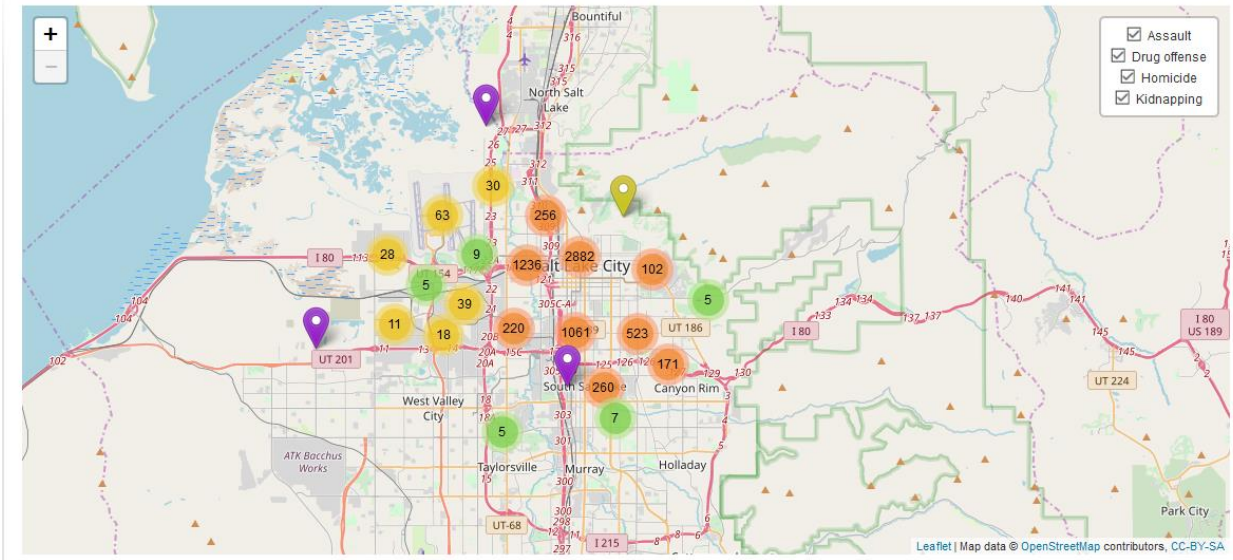
As a first attempt, we were trying to plot all the crimes using small circles with varying opacity. But, since the density of the crimes is so high, we ended up with our map fully crowded with Data points. Also, getting relevant information from this was very difficult.

Second, since the above design wasn't working the way we wanted so we tried using heat maps for representing the crimes. It's a good visualization but the problem here is, we can represent only one crime on the map if we are using a Heat Map. Thus, if the user wants to compare multiple crimes in the same year, it's not possible in this method. We scrapped this idea.

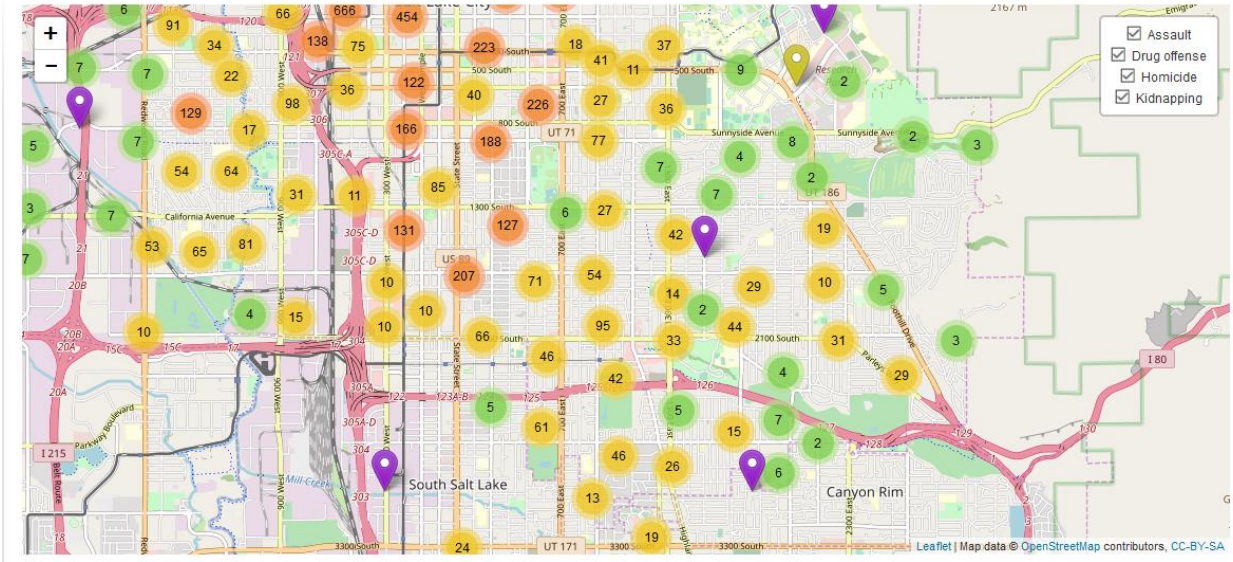
Third, we shifted to the map design which uses Markers to represent the crimes according to the Latitude and Longitude information. Here, we also employed different color markers to differentiate the crime types. It was working fine, and we used robust leaflet libraries to draw markers on the map. Although, the main problem here is, since the data is so large in terms of rows that in some cases there will be too many markers on the map and that would cover the entire map. For an instance, this is the screenshot of a map containing merely 25-30 points and it's already very crowded. 60000 points on this map is not an effective solution.



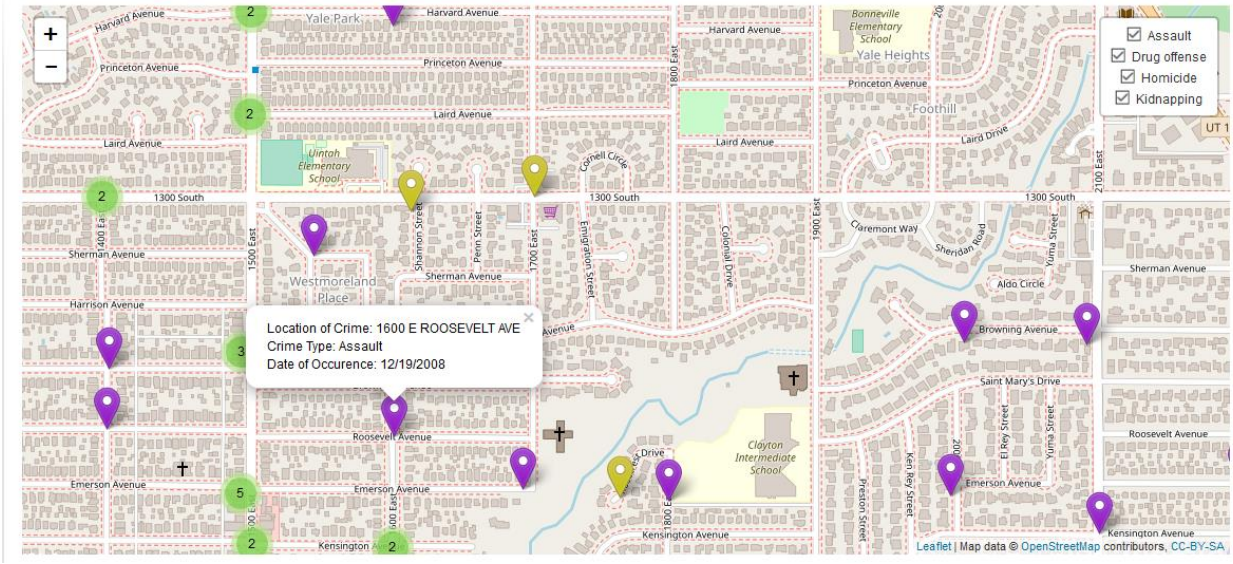
To rectify the above problem, we went with aggregation and clustering of the data points on the map and add Semantic Zooming to get a clean map view. Here, the visualization will start with a high-level overview of the number of crimes (in form of bubbles) in different areas of Salt Lake City and then the user can zoom in to the map and it will unfold the number of crimes into smaller bubbles and individual markers. A user can also click on any of the areas to automatically zoom in to that area. Here, color of the cluster is based on the density of the crimes present in the region. Dark orange means high density and light green means low density. It's a linear scale.



Overview of the map with clustered data points

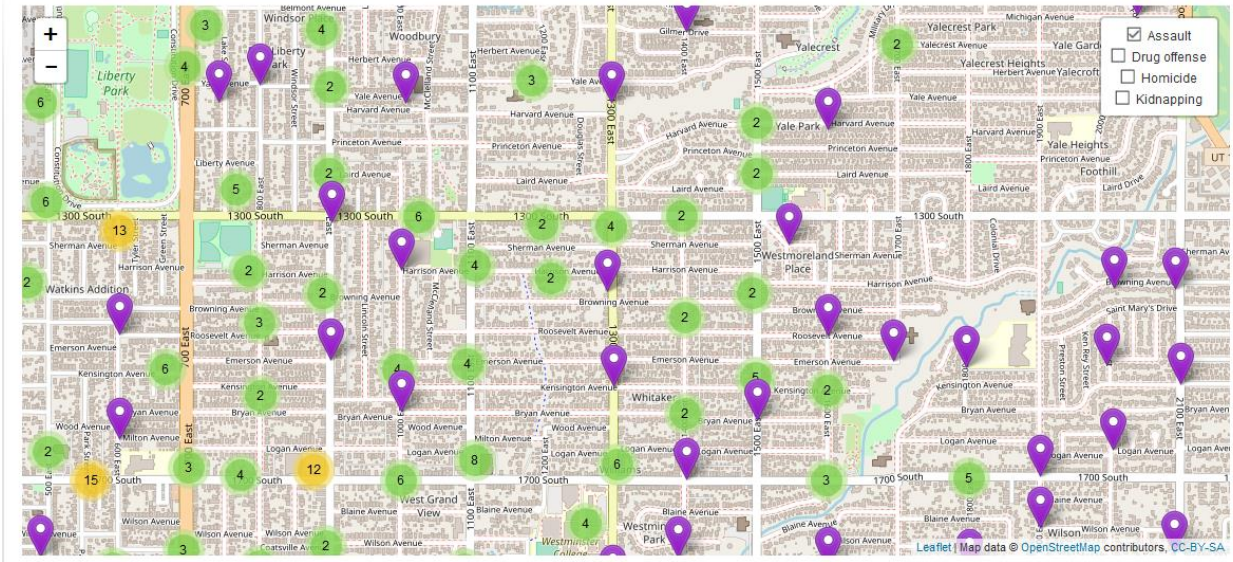


Zooming in the map unfolds the bubbles

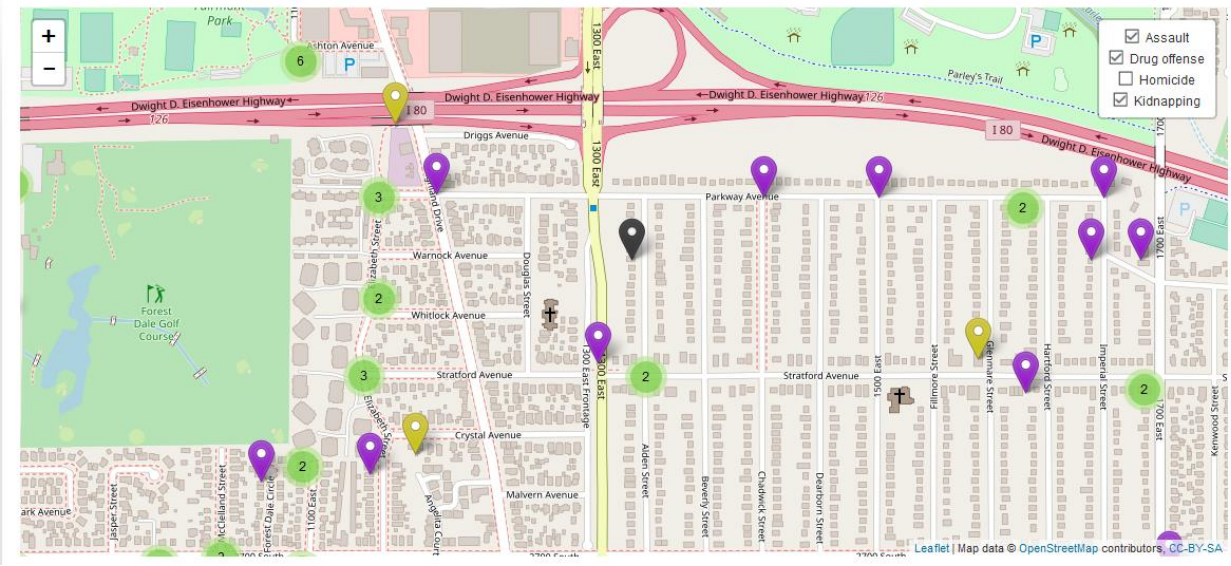


Clicking on a specific popup reveals important info about the crime

Another addition we did here was to create the groups of markers according to the selected crimes. For example, a user selects “Assault” and “Traffic” through the selection view. Markers for those will appear on the map. A user can then choose from an overlay in the map to show only group of Assault or Traffic markers. The checkboxes in the overlay box are exactly same as your current selection from the selection view. It gets updated if the Selection View is updated. An illustration of that is shown in the following images below:

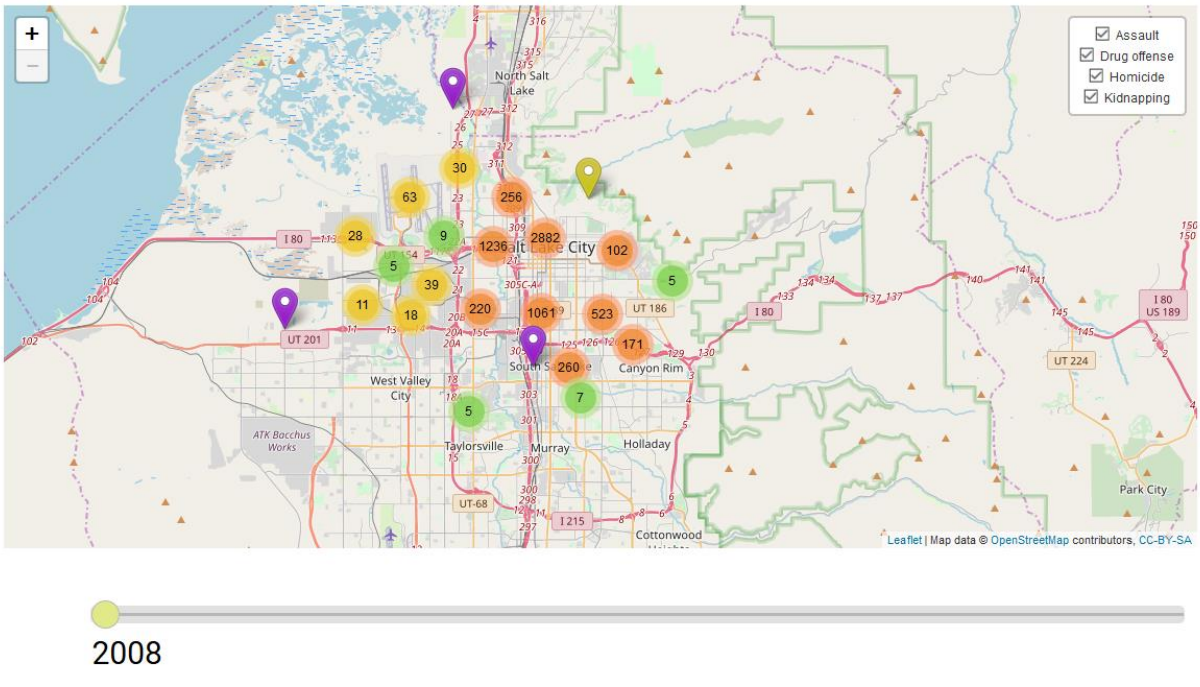


Showing only Assault crimes for the particular year. In the overlay selection, only “Assault” is selected

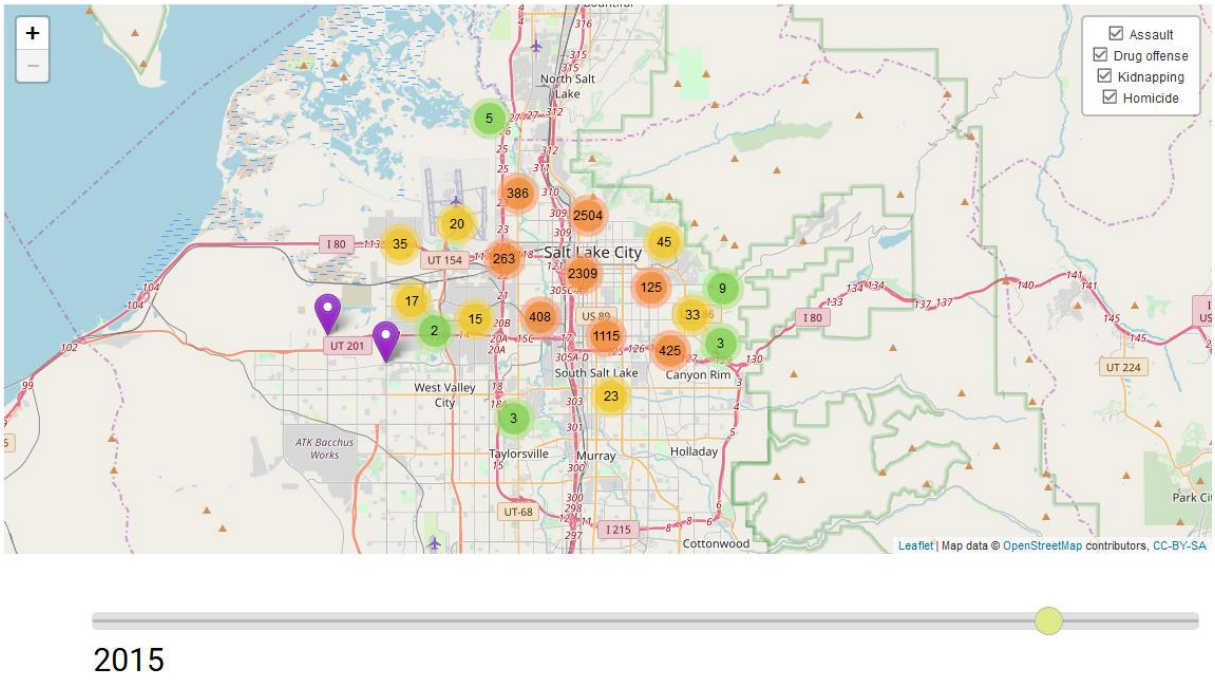


Showing multiple crimes selected with differently colored Markers. Legend is defined in other view.

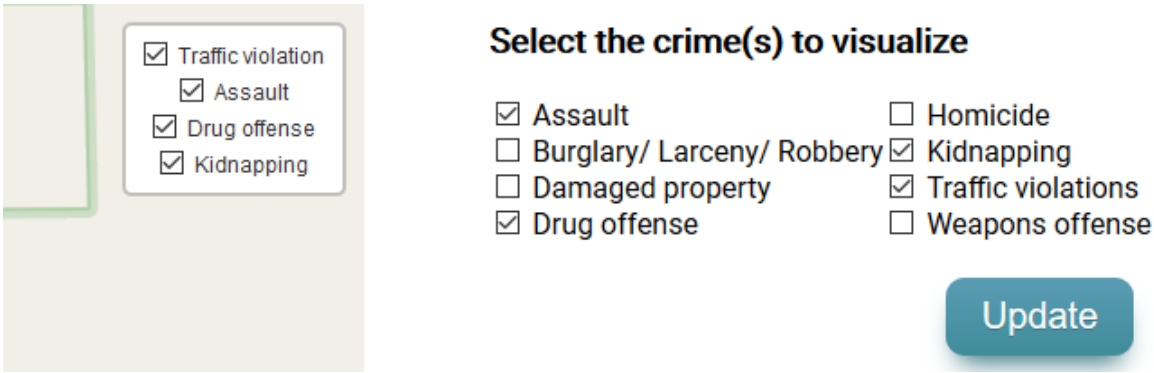
Also, if we change the year using the year slider, map view will get updated to show you the new set of markers based on the new year data and the crime selection. It will retain the crime selection from the crime selection list over the years. But we are updating the overlay box to show all the selected crime types when the year is changed. Examples are shown in the coming images.



Crime Markers for the year 2008



Crime Markers for the year 2015



List of items in the overlay box inside Map is same as the user selection

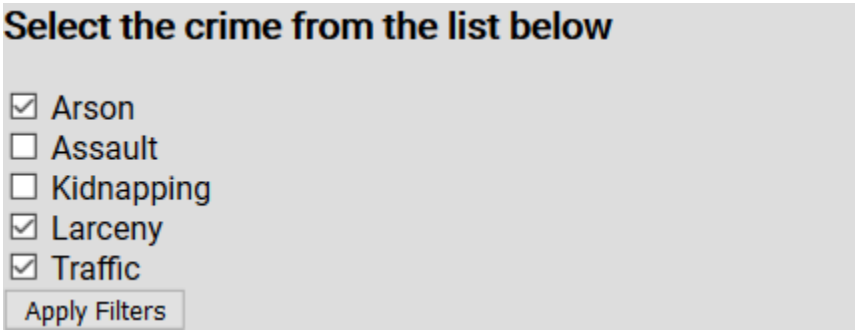
We are using groups for each crime type and adding markers to a crime type corresponding to the selected Crime types. Thus, it provides faster selection when a subgroup is wanting to be selected. For example, in the above examples, first we compute and place the markers for all the selected crime types in the Selection View. Then if the user wants to view a few crime types, he can simply select/deselect the option from the overlay. The map then changes immediately to show those types. Thus, loading all the markers initially helps the system to be faster for the additional queries which will be done later.

Implementation of Map View:

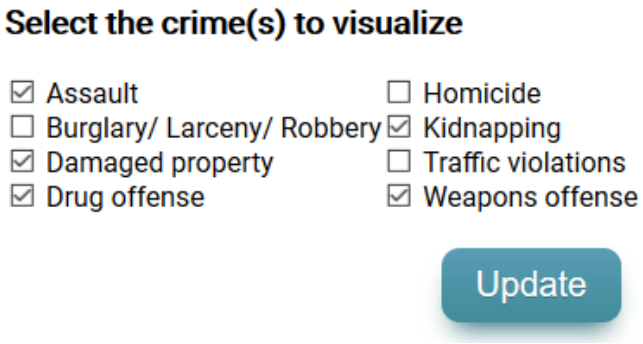
For implementing the Map View, we used *Leaflet version 0.7* as the main tool. For implementing the markers and the marker clusters, we used and modified the internal Marker Cluster Group plugins. Also, to add the overlay layers, we used the in-built functions and plugins like Control Layers, addOverlay etc. Finally, for the markers, we downloaded a set of icons and modified them to work with semantic zooming and the map view. Also, we bind a popup to each of the markers to reflect important information (when clicked) like location of crime, date of occurrence, and type of crime.

3) Selection View

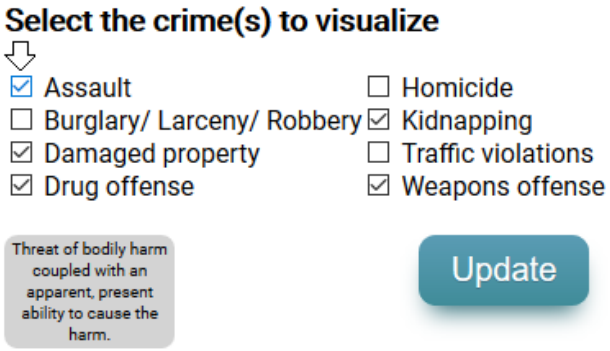
We proposed selection view as a list of checkboxes which are classified based on the crime types. At the time of milestone, we had a selection view something like shown below.



The idea of the selection view is to select crime types from a list of crimes and then analyze how they changed in a year or over the years. By processing the data and aggregating the crimes based on their description, we are now supporting 8 different crime types. Hovering on the checkbox show you a description of that crime type. Finally, clicking on the “Update” button will push the changes to the map and you will get a new set of markers and overlay box depending on the current selection. Also, the Statistics view will get updated based on the new selection. The images for the working are defined below:

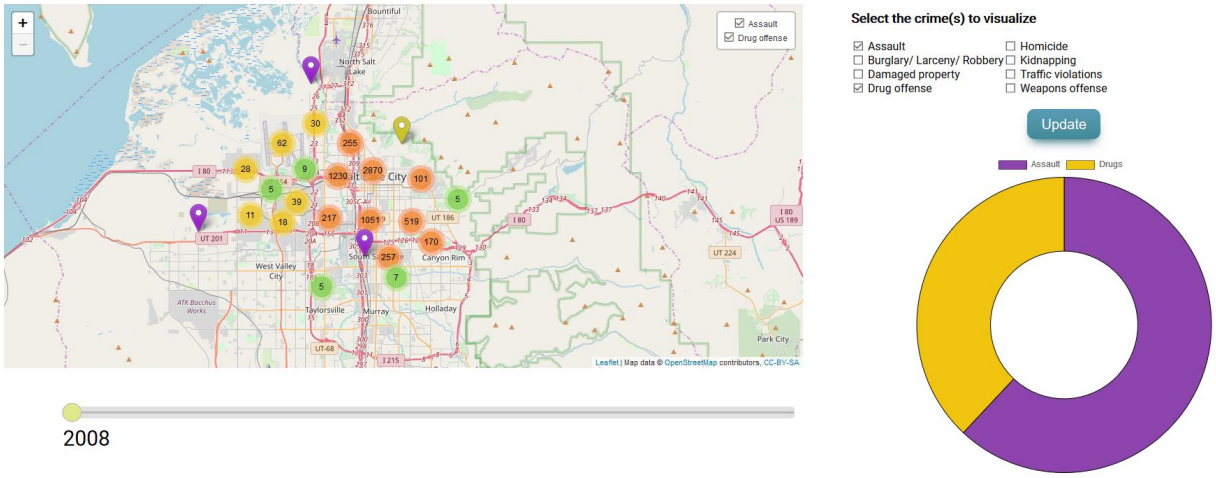


Hovering over any of the checkboxes besides the crime will give you detailed information about the crime in form of a popup. We didn’t link the hovering to the text because usually people don’t want to read the detailed description and we want to keep our selection view.

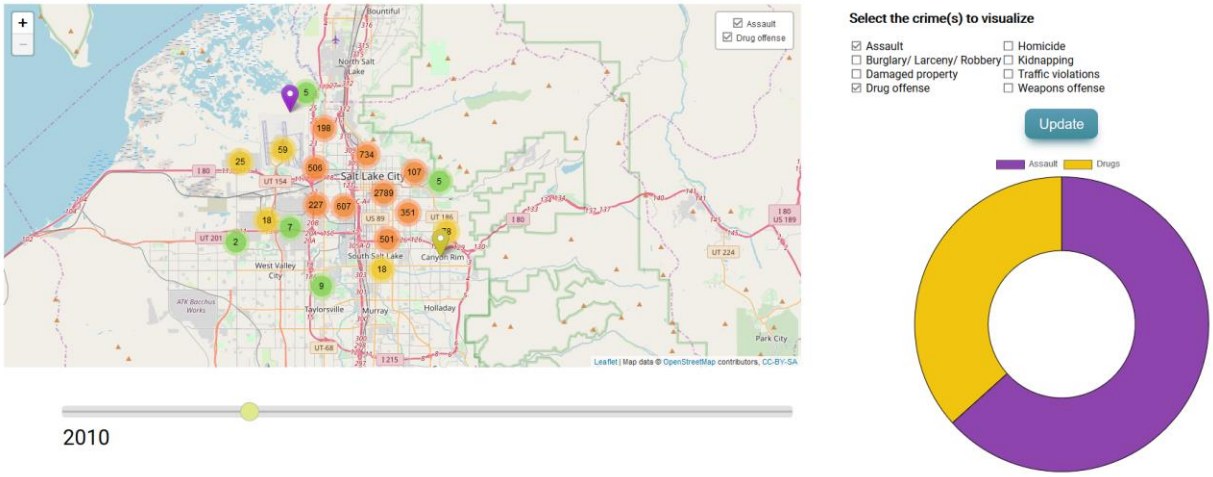


Hovering over "Assault" shows you details about the crime (Black arrow signifies the mouse pointer here)

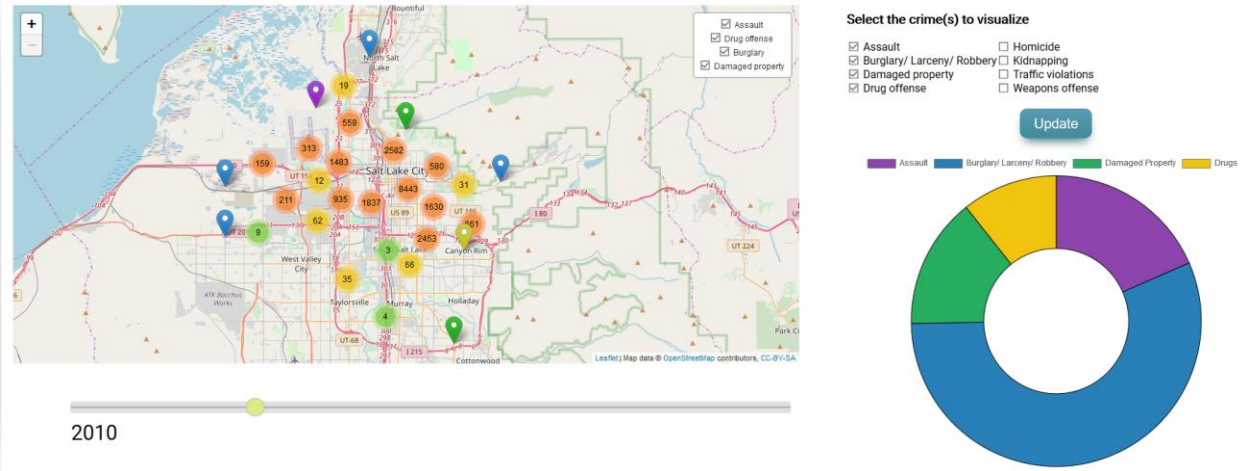
The idea of the selection view is to select crime types from a list of crimes and then analyze how they changed in a year or over the years. Examples below show how the crime selection works;



Default loading of the page includes "Assault" and "Drug Offense". The year is currently 2008 and the map reflects all the crimes from these two types.



Year changed to 2010 with the selection view not changed



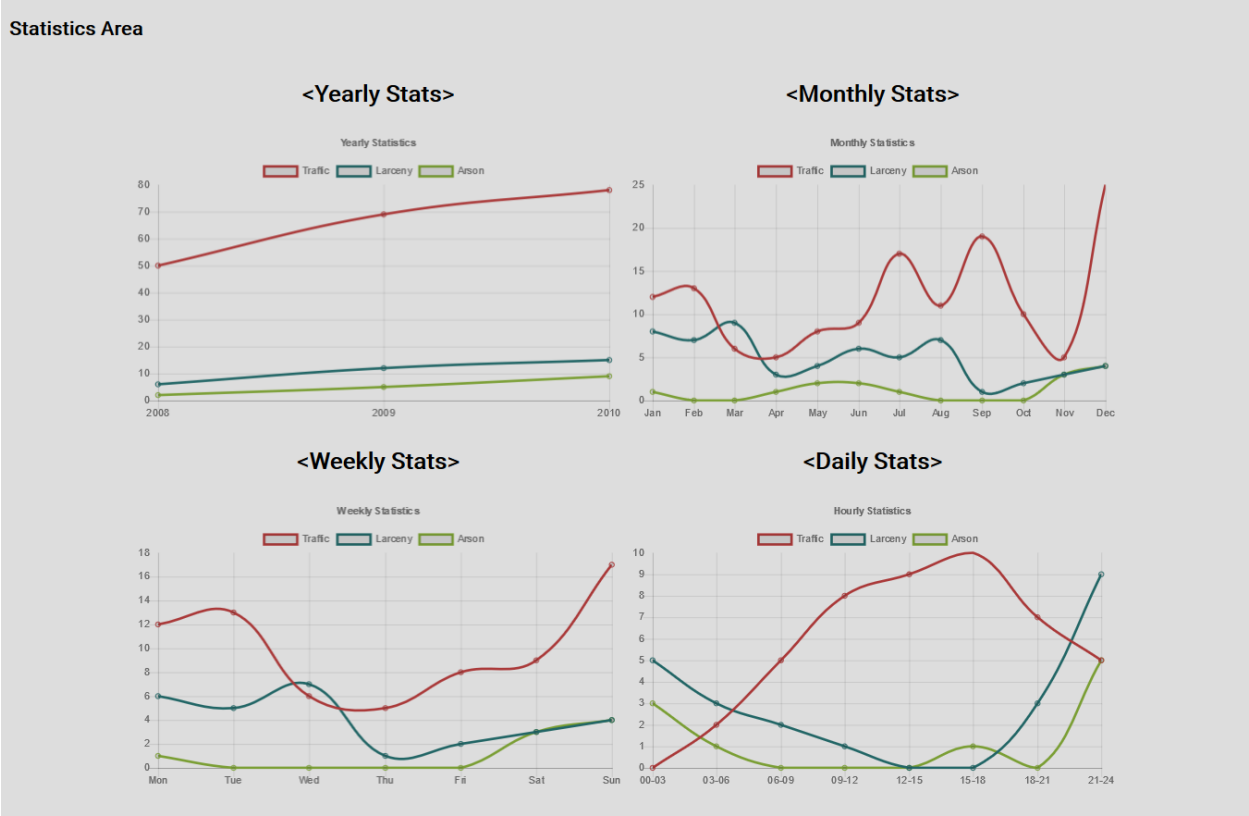
Selection view changed and "Burglary" & "Damaged Property" included in the results

Implementation of Selection View:

We implemented the checkboxes in the selection view simply using a div and multiple input options and the "Update" is a html button. To get the current selection we are using JavaScript event listeners on the checkboxes and on the update button. So, whenever a checkbox is checked and then update is clicked, we call a function "getData()" (we wrote this) and it returns us the current selection of the crimes. We use this list of crimes in the Map View and Statistics View classes to update the views.

4) **Statistics View**

We proposed statistics view as a four aggregated Line Charts which are based on yearly, monthly, daily, and hourly statistics from the data. At the time of milestone, we had a statistics view something like shown below.

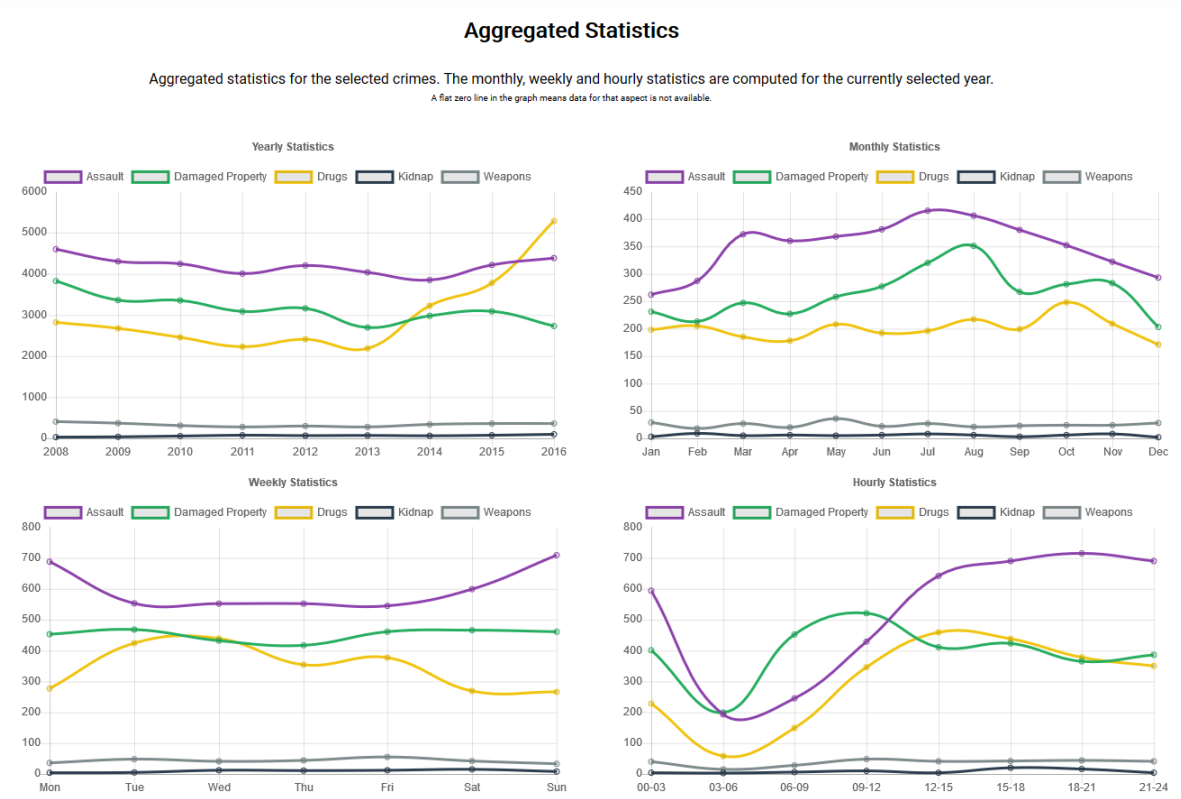


We modified the design and working of this view a bit. Here we are getting the data from the JSON file and the data loading is much faster than before. The legend here matches with the map as well.

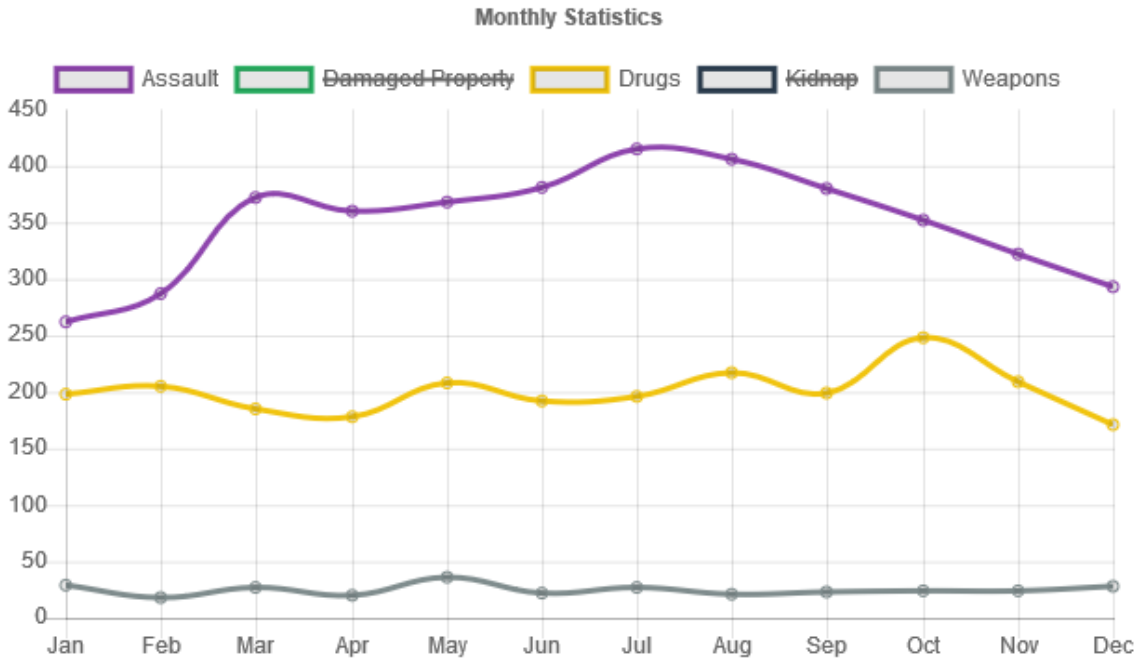
The table on the next page represents the color assigned to each color for the statistics view and the corresponding marker which is consistent in the map view.

Crime Type	Color in the Statistics View	Marker color in the Map
Assault		
Burglary/ Larceny/ Robbery		
Damaged Property		
Drugs		
Homicide		
Traffic violation		
Weapons		
Kidnap		

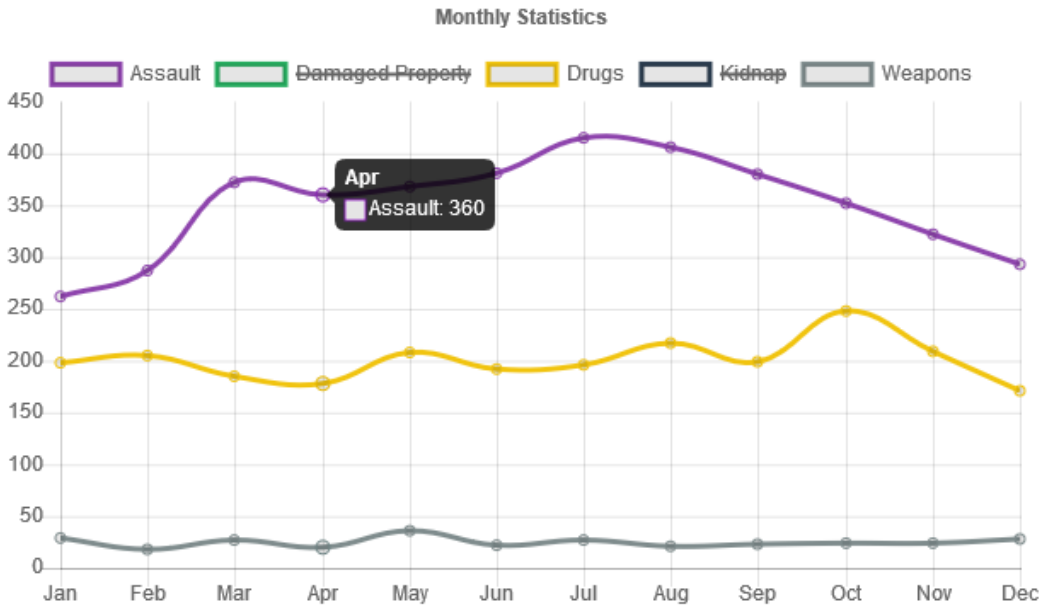
The new statistics view contains 4 charts for yearly, monthly, daily, and hourly data. These are aggregated data statistics for the selected crime. You can check and uncheck crimes in any chart as well. We haven't linked the 4 charts intentionally i.e. if you uncheck a crime in any one chart, it won't reflect the changes in other chart. We did not want to disturb the user's view and distract him with other information and animations. Also, in the hourly chart, we have divided the data in 3-hour windows from 00 to 24 hrs. For the animation, Hovering over Example of working is shown in the images below;



Aggregated Statistics for a particular year with 5 crimes selected



A crime can be checked and unchecked in any of the statistical chart



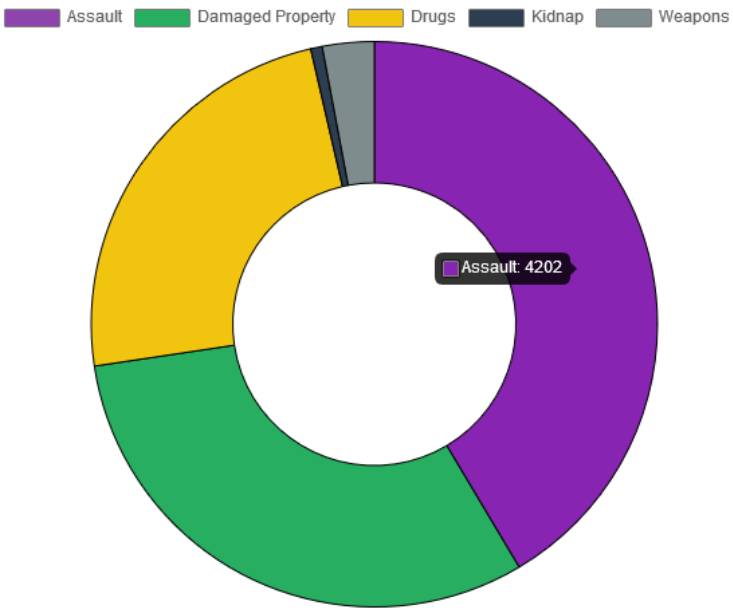
Hovering over any of the "dots" will show you the number of crimes for the particular x value

Implementation of Statistics View

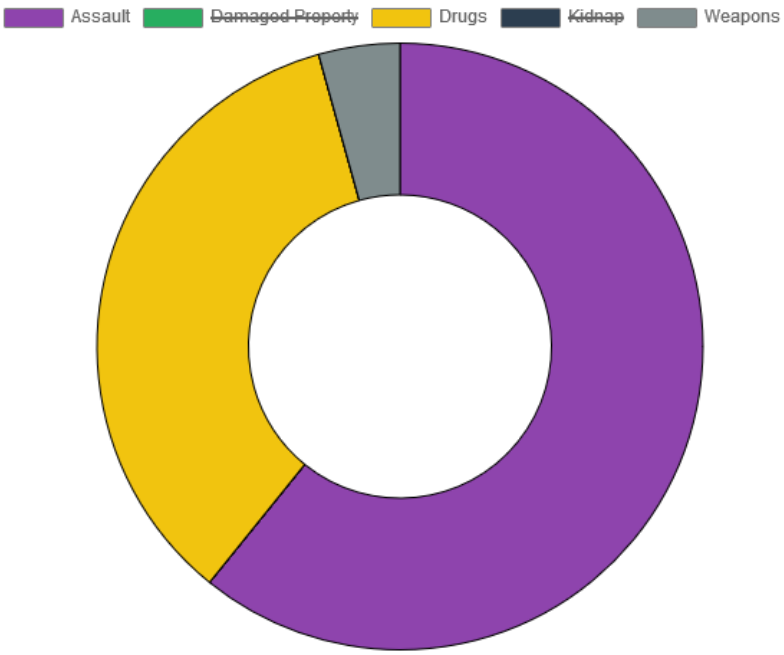
For the implementation of the Statistics view, we used Chart.js as the primary tool. For the data part, we were earlier using the csv files for each year but then we realized that, processing csv takes time if the file contains large number of data rows (~60K per year in our case). Thus, we processed this data and converted it to JSON file. In the final implementation, JSON file is being loaded and passed to different functions in Chart.js. These in-built functions in the library assists us in creating the line charts for the Statistics view and the pie chart for the summary view.

5) Summary View

We also added one additional Summary view which summarizes the statistics for different crime selections and give us an idea about the relative percentage of crimes with respect to each other. Essentially, you go to a year and then select a set of crimes and we show a doughnut chart of the values for those crimes for that year. Hovering over a sector shows you the number of instances for that crime. Images shown below demonstrates the idea;



Summary view for the year 2012 for the selected crimes



Similar to the Statistics View, a user can check/uncheck a particular crime

Implementation of Summary View:

For the summary view, we used Chart.js. We are passing the data for the selected crimes and selected year to the in-built function which creates the Pie chart out of the data.

QUESTIONS

Many questions arise when we were creating the visualizations and processing the data. Some of them are listed below;

- What's the general crime landscape in Salt Lake City and how it has changed over the years? Essentially, an overview about the living in Salt Lake City.
- Are there any crimes that have been constantly increasing over time?
- If a person wants to buy a house, he/she can query what kind of crimes are prevalent near the location.
- How's one neighborhood different from other neighborhoods in terms of crime types?
- What hour of the day/night is the most dangerous in terms of crimes in a neighborhood? Ex. what's the number of stealing incidents on Friday or Saturday night etc.

EVALUATIONS

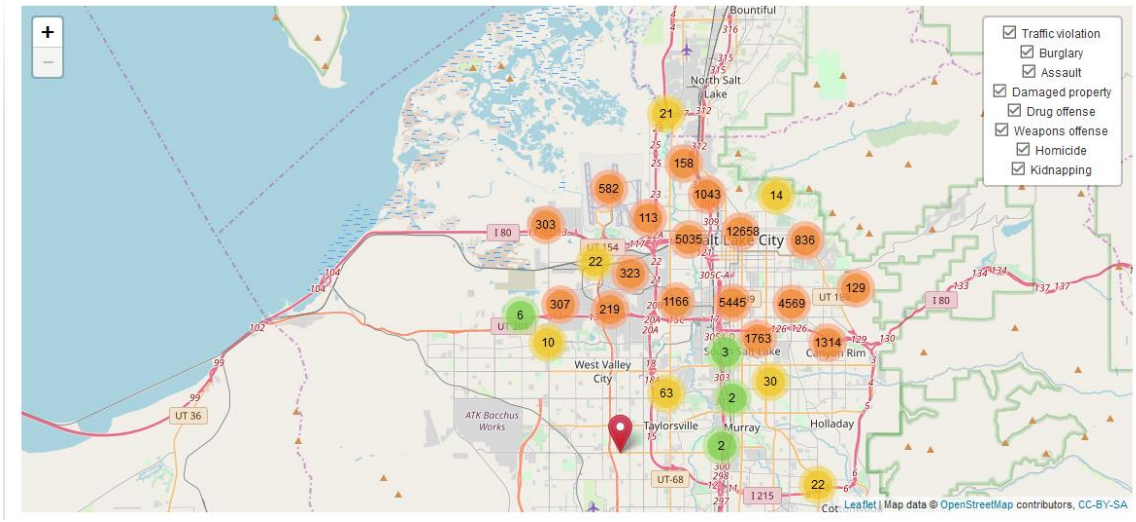
Our visualization is meant to be used as a utility tool using which the users can use to get acquainted to the Crime Landscape of Salt Lake City. This tool is developed such that it can be used by anyone with or without any knowledge of Visualization. Target audience for this project are current residents of Salt Lake City, people who may visit in Salt Lake City or people who plan to move in Salt Lake City in the near future.

The Map View and the Statistics View enable us to understand the crime landscape of Salt Lake City from different perspectives. One can use this tool to answer several queries based on crimes over the years. Some examples of questions that can be answered from specific views are shown below.

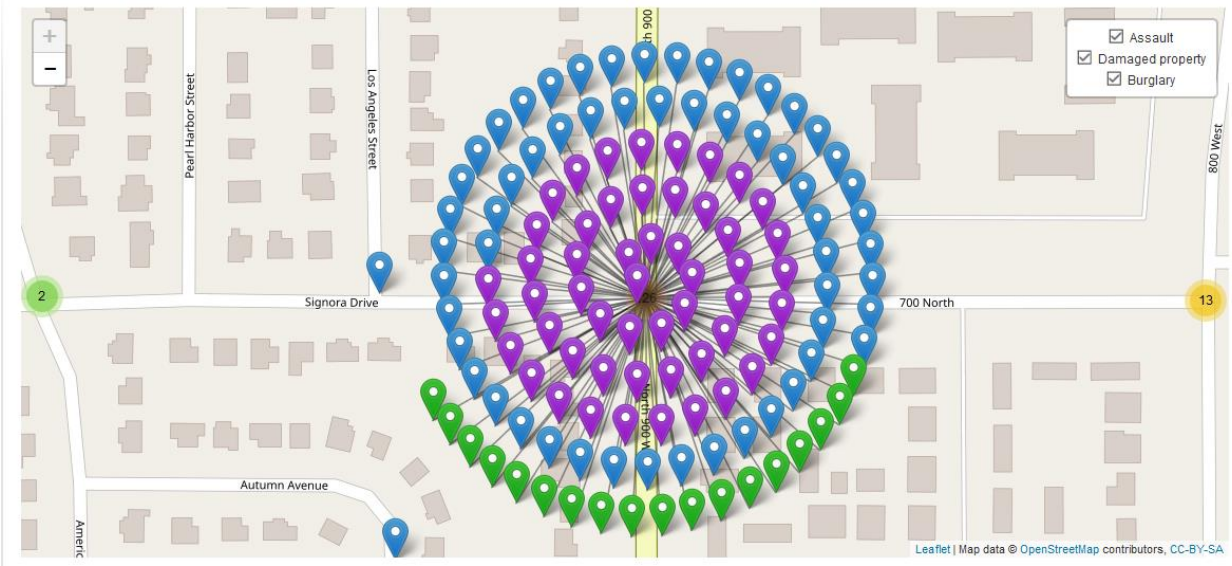
Map View

The Map View shows the distribution of crimes in different regions of Salt Lake City over the years giving us a geographical landscape of the crimes.

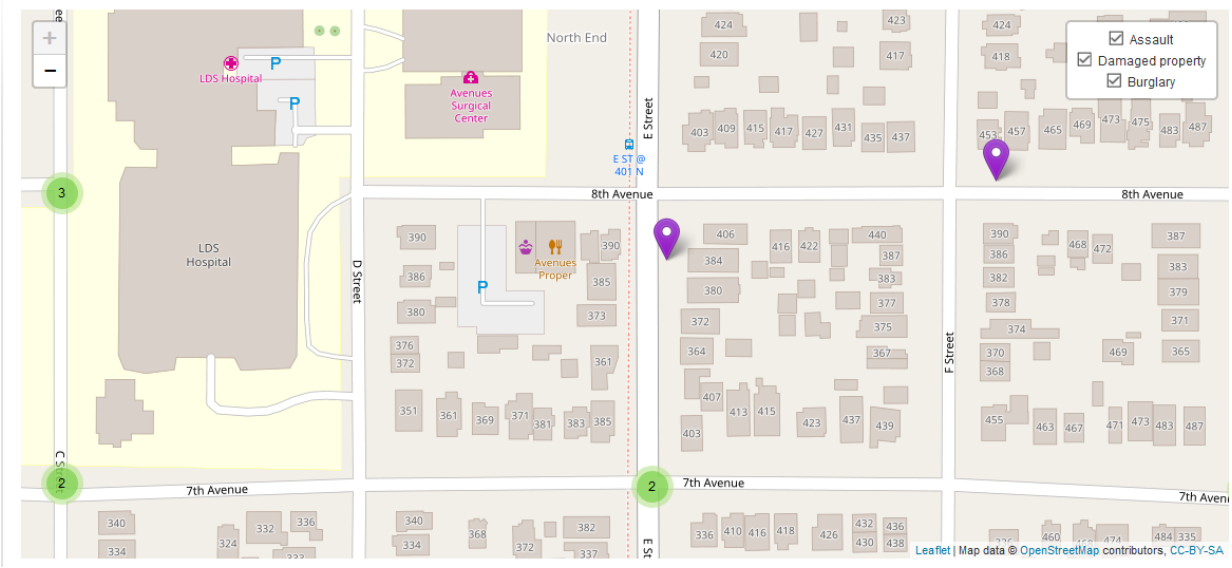
- We found that most of the crimes are more concentrated in the city compared to the outskirts. Following is the entire crime landscape for the year 2008.



- If a person is looking for a residence in Salt Lake City, they can use this tool to learn more about the crime trends in residential neighborhood. We selected some crimes relevant to residential areas like Damaged Property, Burglary and Assault. We used this to find some good neighborhoods and some troublesome neighborhoods.



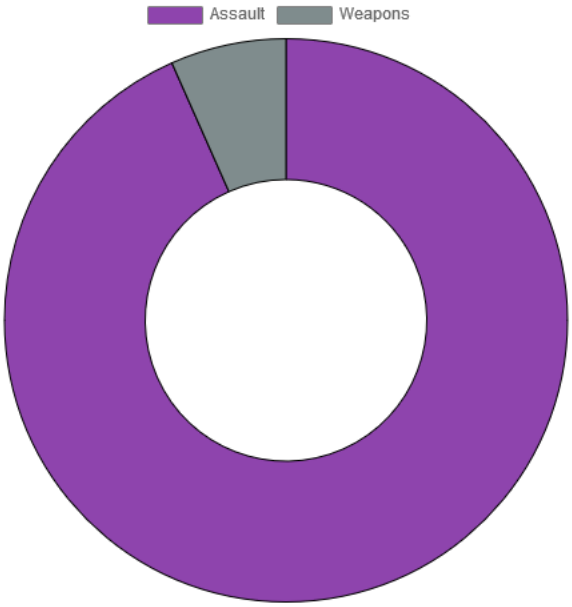
For an instance, this particular neighborhood 700 S 900 W has been prone to residential crimes in the past



On the other hand, this particular neighborhood in the Avenues has seen minimal number of residential crimes.

Summary View

The Summary View shows the percentage of crimes for any selected year. From this view one can compare individual crimes by number of occurrences in that year.

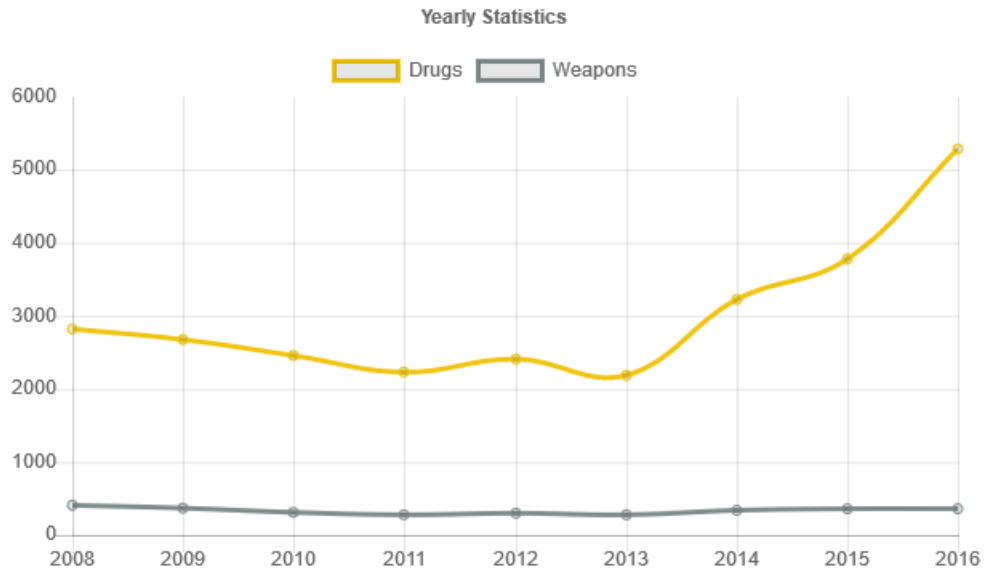


This view shows us that for a given year, number of Weapons offenses is much lower than the number of Assaults.

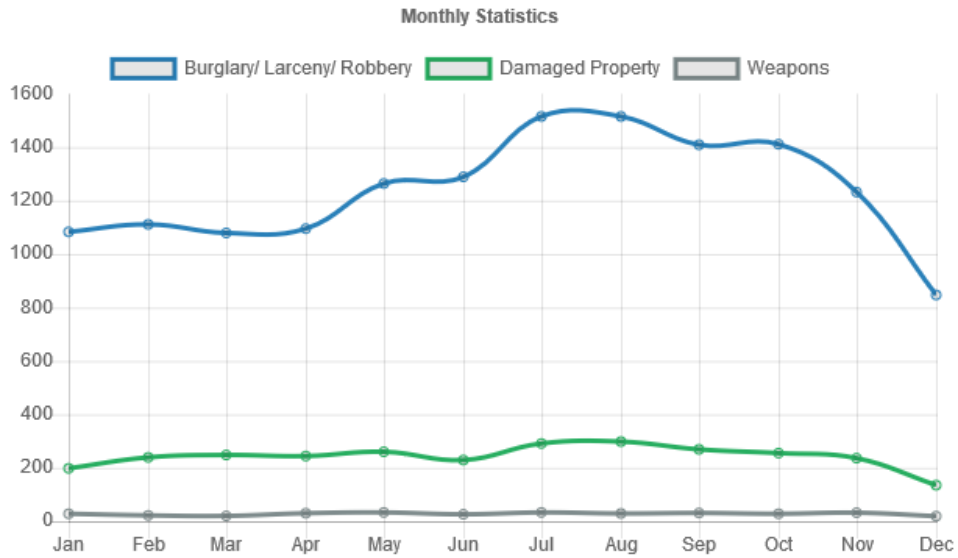
Statistics View

The Statistics View shows the trends of the crimes over the years, month of year, day of week, and time of day.

- Using the Yearly Statistics we found that the usage of drugs has increased over the years and there has been a significant increase in the 2014, 2015 and 2016.

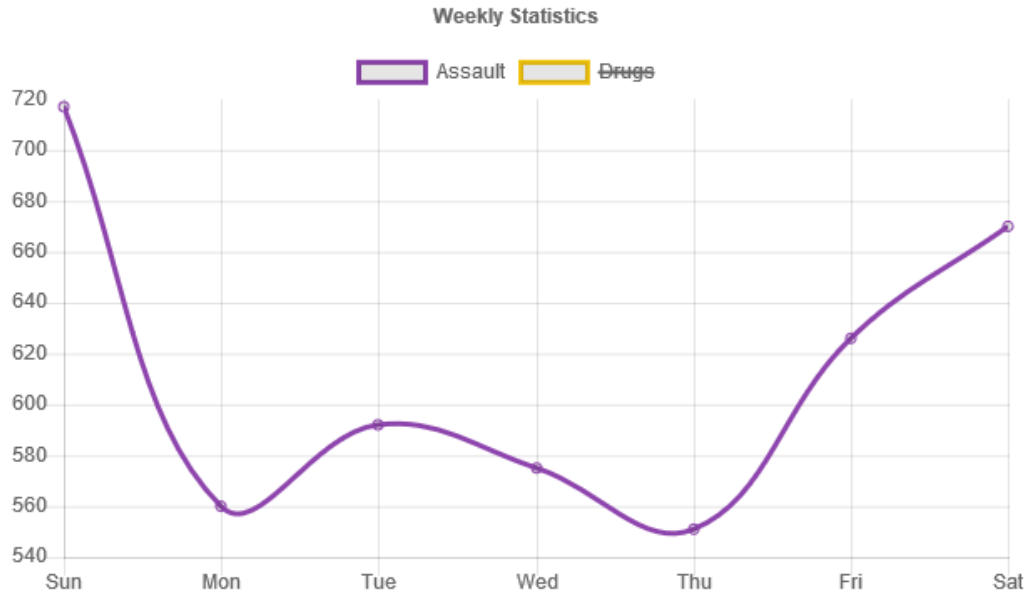


- Using the Monthly Statistics we found that burglary, larceny and robbery decreases during the holidays from the months of October to December.



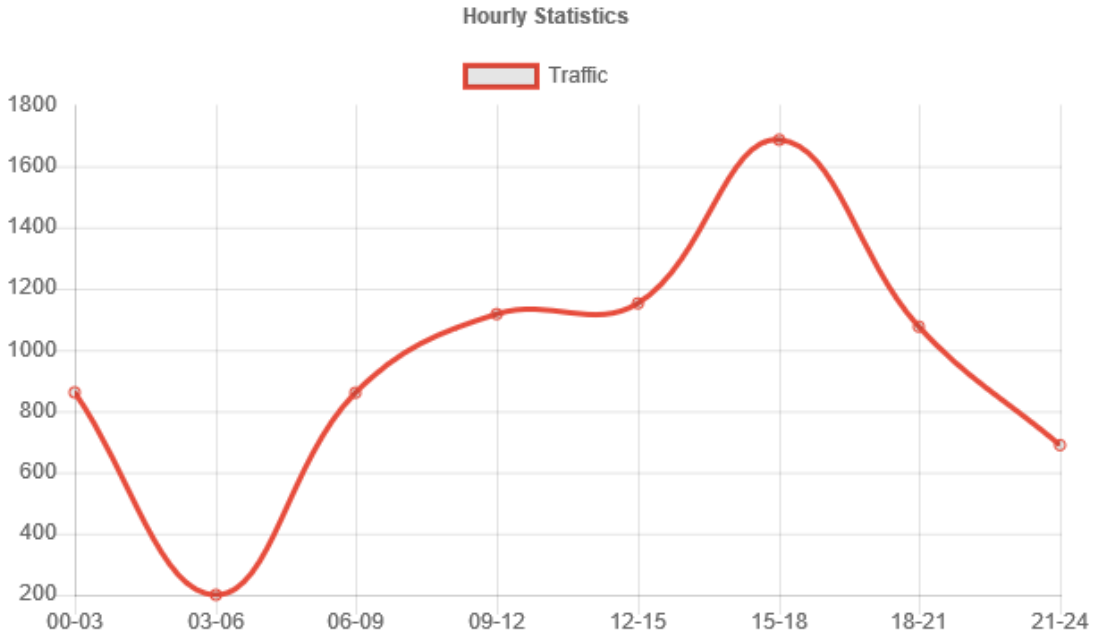
Burglary, Larceny and Robbery rates decreases a lot in holidays contrary to the idea of more thefts happening in Holidays

- Using the Weekly Statistics we found that assault rates are high during the weekend



Assaults can be observed to be really high on the Weekends. The reason might be because these are non-working days and people roam around in the city till late night.

- Using the Hourly Statistics we found that traffic violations are at the peak between 3 pm and 6pm.

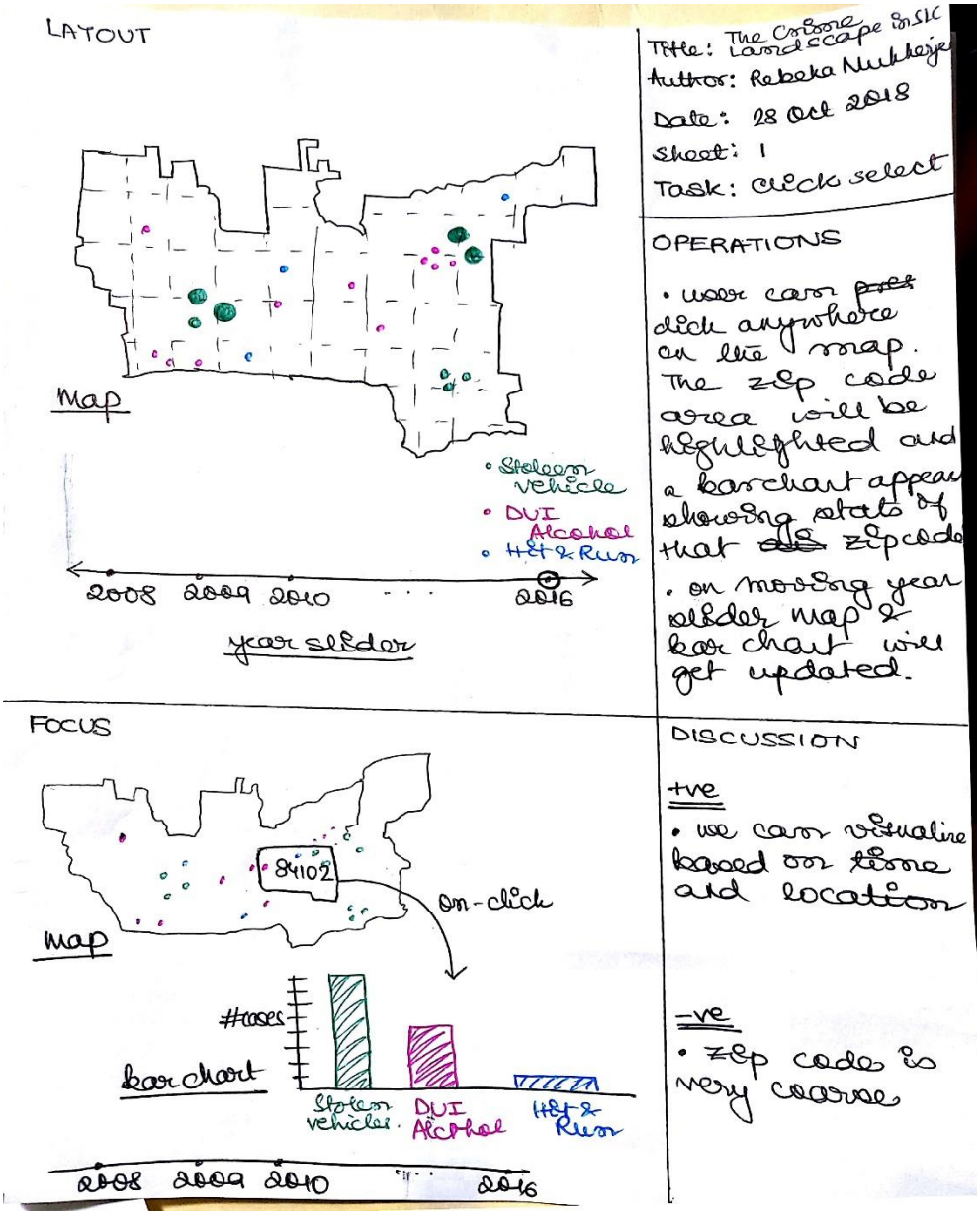


For ex. in year 2014, the traffic violations were at peak between 3 - 6 PM. Follows the general idea of work hours in the evening.

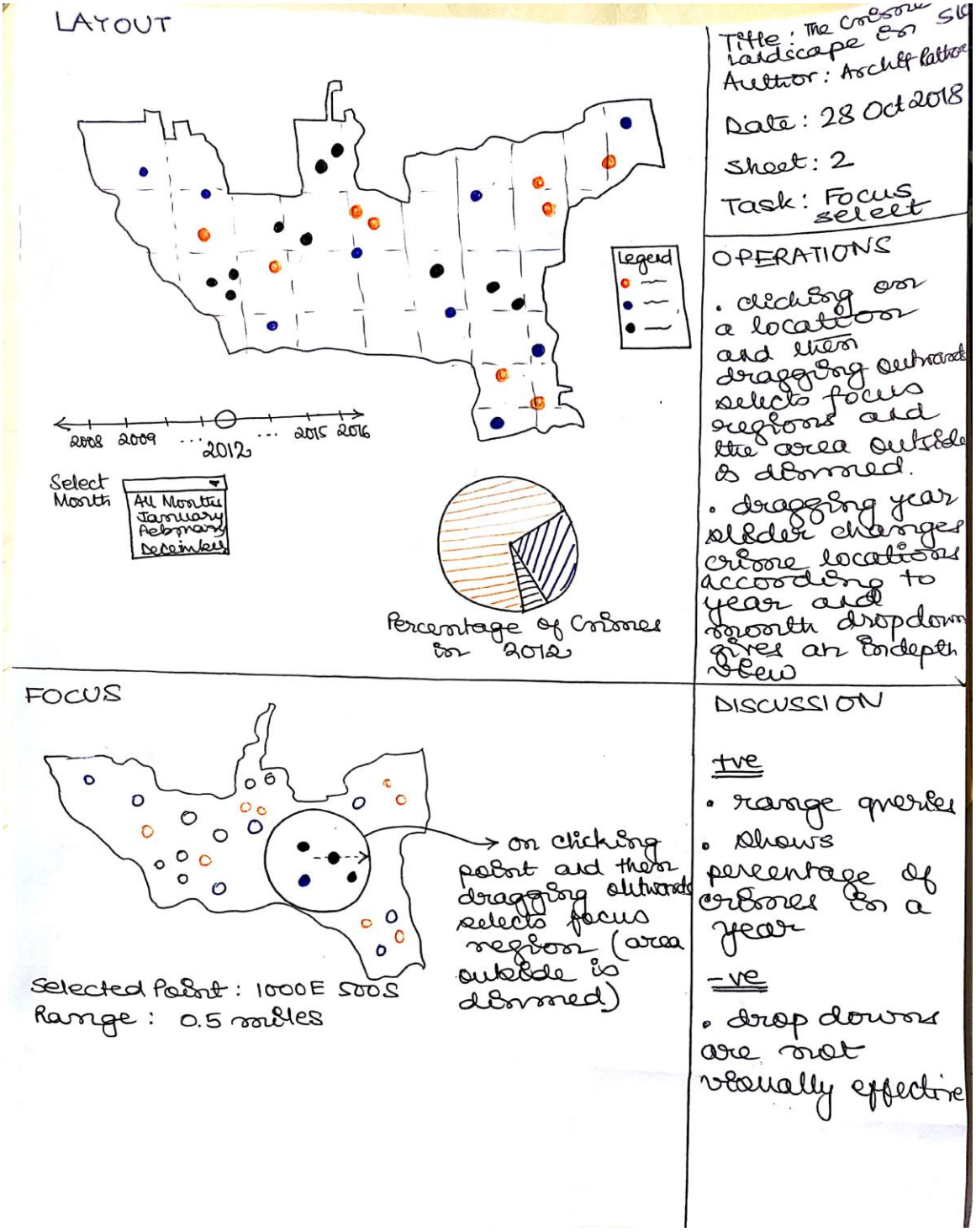
REFERENCES

- https://opendata.utah.gov/
- https://leafletjs.com/reference-1.3.4.html
- https://www.chartjs.org/docs/latest/
- https://github.com/Leaflet/Leaflet.markercluster
- Wikipedia and online legal dictionaries for definitions of crimes

APPENDIX – PROPOSAL IMAGES

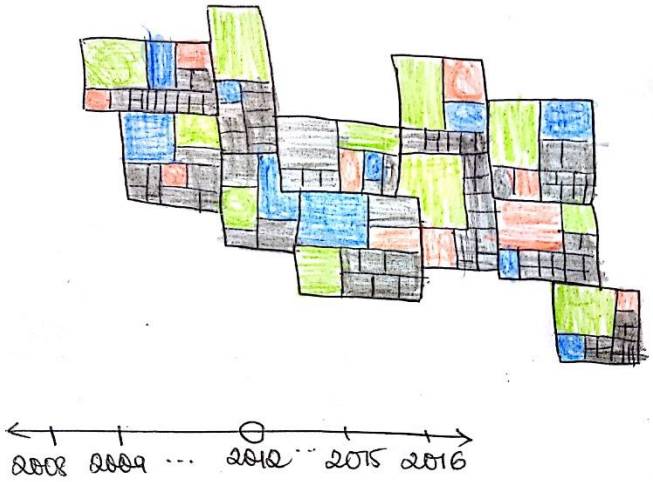


Prototype 1 Design



Prototype 2 Design

LAYOUT

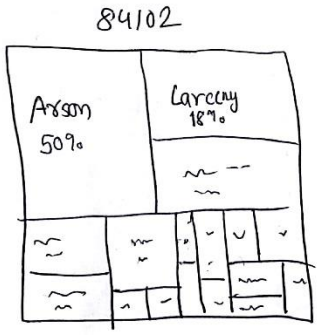


Title: The Crime Landscape in SLC
Author: Yash Gargade
Date: 28 Oct 2018
Sheet: 3
Task: Treemap

OPERATIONS

- Crimes are shown as a treemap based on zipcodes
- On clicking on zip code area the percentage of each crime is displayed

FOCUS



DISCUSSION

+ve

- the treemap visually encode crimes percentage by location

-ve

- the treemap is not visually effective.

Prototype 3 Design