Crimes In Baltimore 2012-2017

Process Book

Group No: 6

Name: Bhushan Marathe

Email: bmarath@clemson.edu

Clemson id: C16217536

Name: Shivam Nahar

Email: snahar@clemson.edu

Clemson Id: C18487863

Name: Hemath Reddy Dasari

Emil: hdasari@clemson.edu

Clemson id: C1058679

Video Link: https://onedrive.live.com/?authkey=%21AD3LgH83YJK9XYo&id=D9F178DD54E013%21718&cid=00D9F178DD54E013&parId=root&parQt=sharedby&parCid=18BC808AA732A347&o=OneUp

Website Link: https://imsherlocked.github.io/DataVisualization_Project_6030/

Repository Link: imsherlocked/DataVisualization_Project_6030 (github.com)

Dataset Link: https://crime-data-explorer.app.cloud.gov/pages/home

1 Overview: Our main overview for our current project was to show the bar chart, based upon the intensity and impact which will help identity the locations[neighborhood] which has the highest crimes. We also wanted to show another visualization to identify the hotspots based on the latitude and longitudes.

1.1 Motivation: Nowadays the crime rates have significantly increased than the previous years. The security and safety have become a necessity for which the government have invested millions of dollars so that these crimes can be reduced. The crime data has been captured and is monitored to compare the rise or fall of such criminal activities. The same has been done by the state of Baltimore Police Department in which the criminal activities were captured from 2012-2017. Our motivation for selecting this dataset: **"Crimes in Baltimore"** will help analyze the data and create visualization which will help the end user to get the data just by looking at the visualization, rather than spending time on the verbose documents. For this, we referred multiple research papers such as CriPAV: Street-Level Crime Patterns Analysis and Visualization [1], Analysis and Visualization of City Crimes [2], A review of Crime Analysis and Visualization. Case study: Maryland State, USA [3].

2 Related Work: To start with our journey, we went through multiple YouTube videos and research papers in which we found out some research papers which were related to visualizations for the crime data. From those papers we understood, the crimes which occur at certain location and the frequently occurred crime types which can be shown as a visualization. The first research paper we took as a reference is "A review of Crime Analysis and Visualization. Case Study: Maryland State, USA". This research paper mainly focuses on geo visualization and map visualization. We found this paper interesting as they defined a geographical information system and exploratory data analysis to show case the crimes visualization. The second research paper we took as a reference is "CriPAV: Street-Level Crime Patterns Analysis and Visualization". The journal's executive summary states that a

new tool called CriPAV was developed and is utilized to provide a street-level understanding of crime patterns. This study report primarily focuses on comparing the probability and intensity of crimes. We found this article interesting since it used a different approach to classifying hotspots by comparing probability to intensity. In the third paper they used the latitudes and longitude to plot on the leaflet and show a heatmap based on the crimes frequency for the data received from the Police Department. Taking inspiration from the above papers we wanted to use them in our Baltimore crimes from 2012 to 2017. We used the ideas of scatterplot which will show the intensity and probability of the crimes based on the neighborhood data which will be shown on the leaflet map. The line chart to show the trend of the total crimes for our dataset from 2012-2017 and a geoJSON heatmap to display all the nearby counties and the color hue to show the minimum to maximum crimes mentioned in our dataset in out project.

3 Questions:

Q1. What are the hotspots in Baltimore city based on the probability of the crimes?

Q2. What is the crime that has the highest probability and intensity of occurring in a neighborhood?

Q3. What is the overall trend of crimes over the years?

Q4. Which area/neighborhood with total crime and trend from 2012-2017 crimes?

All these questions evolved gradually and at the initial stage of the project, it seemed to be easy, but as we started developing the visualizations, we must do many things with our data. We have done a lot of data preprocessing and used several libraries in JavaScript. In the end, we got the desired output.

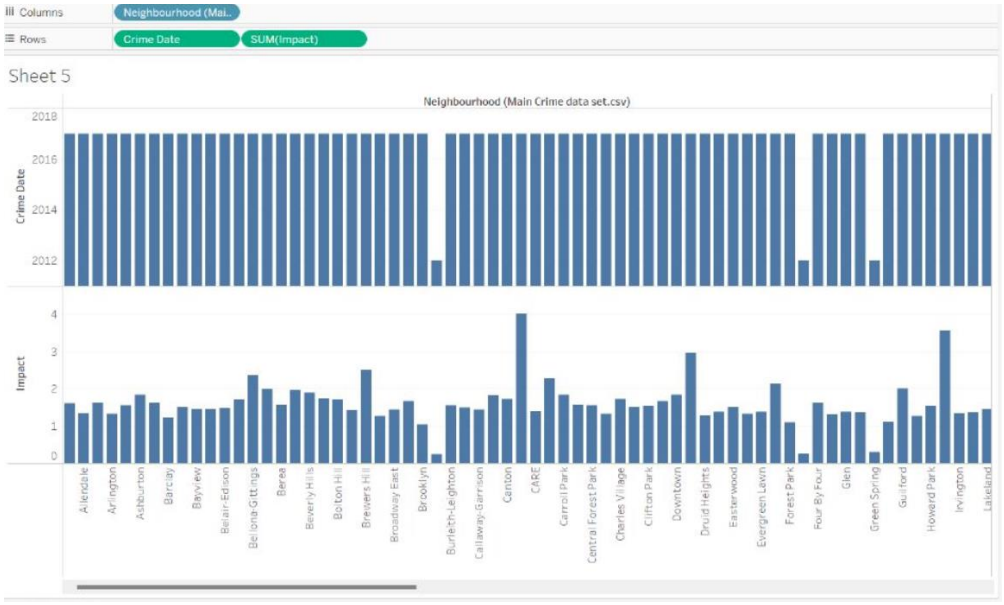
4 Data:

We collected the data from official Government website of Maryland which consists of crime data from 2012 to 2017. There were some null values in the dataset which we removed by data preprocessing. We have considered the crimes which are likely to occur. To reach the goals of our project, we should have the probability and intensity of the crimes that we did not have in the dataset. For this, we calculated the Probability and Intensity of the crimes in each location. The probability is calculated using the formula max(crime)/total number of crimes at a particular location. The crime rate, which is the greatest in each location, the probability that the crime will occur, and the intensity, which is simply the total number of crimes that have occurred at that location, have been taken into consideration. Also, we have used Date Time to remove the day and month from year to present the data

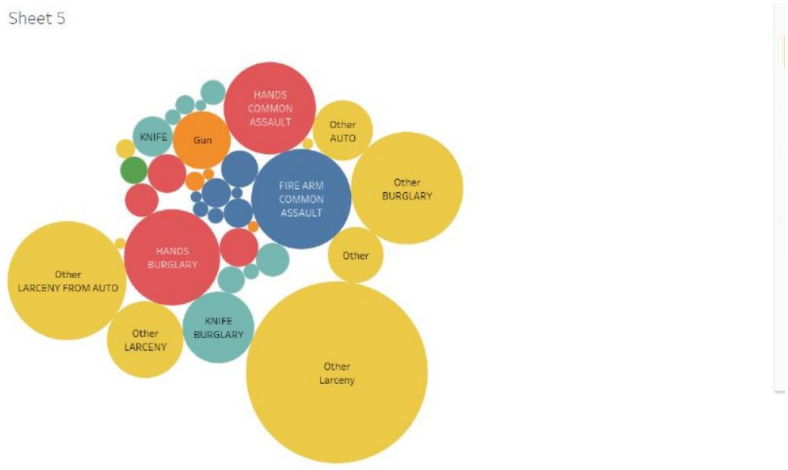
year-wise. Through this we have created six CSV files from 2012 to 2017. After computing the probability and intensity we have merged all six files into one csv file. Finally, we had to prepare some data for our forthcoming encounter. Before creating a dataset using our data, we first downloaded a geojson file Community Statistical Area from the web. According to the geojson information we have, we combined two to three places and totaled all crimes committed in those areas during a six-year period. After this we created our data according to the our dataset by combining the neighborhoods.

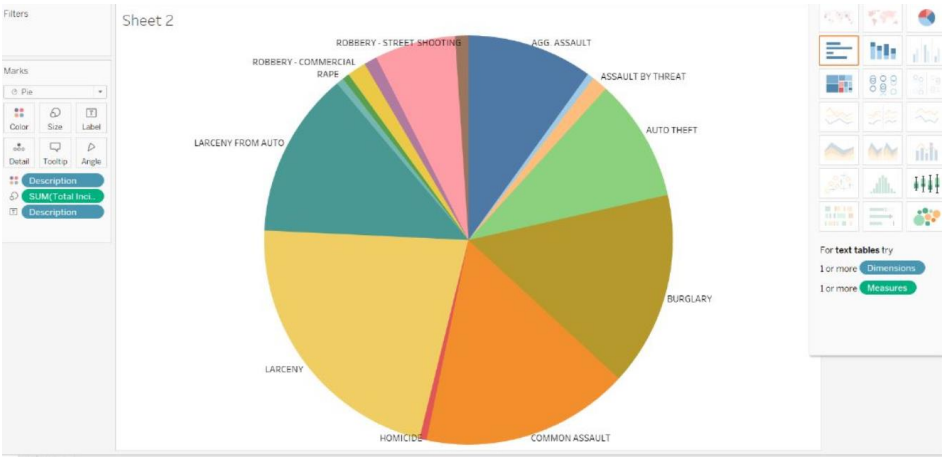
5 Exploratory Data Analysis

: At the starting stage of the project, we could not replicate the probability intensity scatter plot. We started the project with aim to replicate the techniques that we have implemented in tableau to get the same outputs in our project. As we are doing our project on crime hotspots, the intention is to create the scatter plot which will change according to the dropdown selections. But later, Dr. Federico commented that the visualization is too simple from the user point of view, so we have used scatter plot only for the reference. Then we started searching for the best library to create the street level map. We saw the leaflet a JavaScript library and found that good to create a street level map. We were able to use the locations to make a map of Baltimore. Then, we started the visualization component and created duplicate data to make sure we were receiving the appropriate visualizations. Then, we started the visualization component and created duplicate data to make sure we were receiving the appropriate visualizations. First, we have produced a scatter plot and a duplicate CSV file containing probability and intensity. The intended visuals were achieved. Then we have created a duplicate geojson file and used leaflet, an open-source JavaScript library. We placed the crime sites on the map using the coordinates for Baltimore and the latitudes and longitudes from our dataset. Then, using our dataset, we repeated the process, made the map, and plotted the actual locations on it. We also utilized the Observable website as a guide to see if all of our visualizations were coming through correctly. Additionally, we saw various geospatial visualizations on the observable website before utilizing JavaScript to build the map. We received all our results as expected and all the necessary insights. We use these insights up until the project's conclusion to help us produce the results we want.


Columns
Neighbourhood (Mai..
Rows
Crime Date
SUM(Impact)
Sheet 5
Neighbourhood (Main Crime data set.csv)
Crime Date
2018
2016
2014
2012
Impact
4
3
2
1
0
Allendale
Arlington
Ashburton
Barclay
Bayview
Belair-Edison
Beltona-Gittings
Berea
Beverly Hills
Bolton Hill
Brewers Hill
Broadway East
Brooklyn
Burleith-Leighton
Callaway-Garrison
Canton
CARE
Carroll Park
Central Forest Park
Charles Village
Clifton Park
Downtown
Druid Heights
Easterwood
Evergreen-Lawn
Forest Park
Four By Four
Glen
Green Spring
Guilford
Howard Park
Irvington
Lakeland

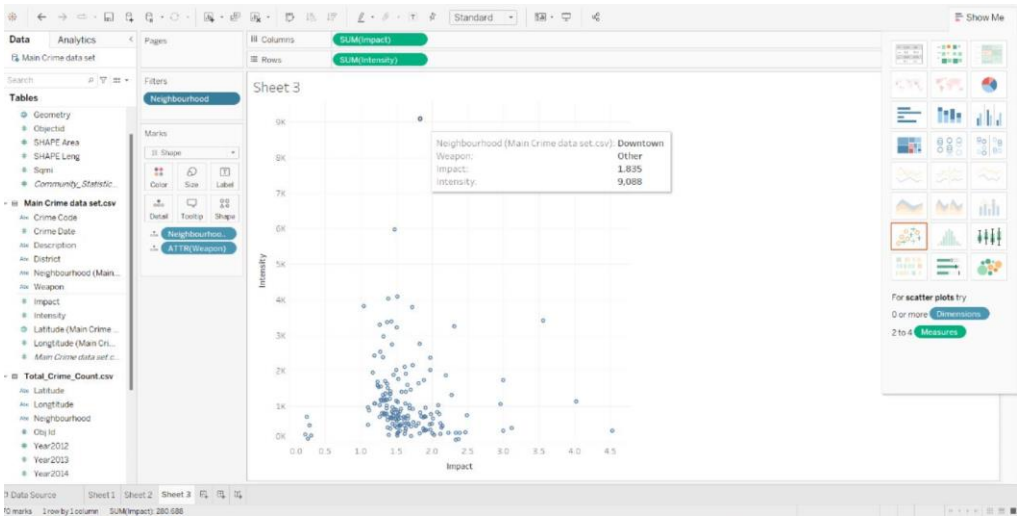

Sheet 5
Other Larceny Port Covington
Other Larceny Canton Industrial
Other Larceny Inner Harbor
Other Larceny Seton
Other
Other Larceny
Other
HANDS
KNIFE
Gun


Sheet 5
HANDS COMMON ASSAULT
Other AUTO
KNIFE
Gun
FIRE ARM COMMON ASSAULT
Other BURGLARY
HANDS BURGLARY
Other
Other LARCENY FROM AUTO
Other LARCENY
KNIFE BURGLARY
Other Larceny



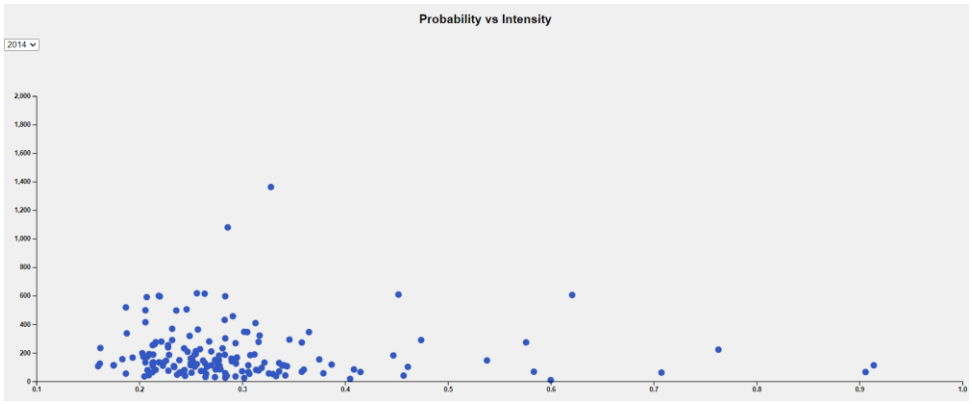
6 Design Evolution: After picking the topic and dataset, we thought of doing the bar chart, Sankey graph, and Scatter plot as our visualizations to present the data. Later Dr. Federico suggested that we need to do interactions to visualize our data so that it will be easy for the user to understand the data. Then we stopped doing the above visualizations. After understanding the dataset, we started computing probability for all the locations from 2012 to 2017. We chose the crime that has occurred the most frequently in a certain area, and then we reasoned that a scatter plot would be the best way to visualize the data by selecting the dependent variable "Intensity." Using the scatter plot for this visualization, we can easily plot the data in a scatter plot. We started using the observable site and then created the Sankey chart. We drew a map of Baltimore City with the aid of a brochure and then marked the hotspots on the map. The dropdown option to choose a year was then retained for choosing the years. At that point, our scatter plot and the geomap both appear fantastic. We shouldn't use the observable site, Dr. Federico said. Additionally, he advised us to only use the observable website as a guide and instructed us to write the code in a JavaScript file. The scatter plot in d3 was next created.

Tableau:

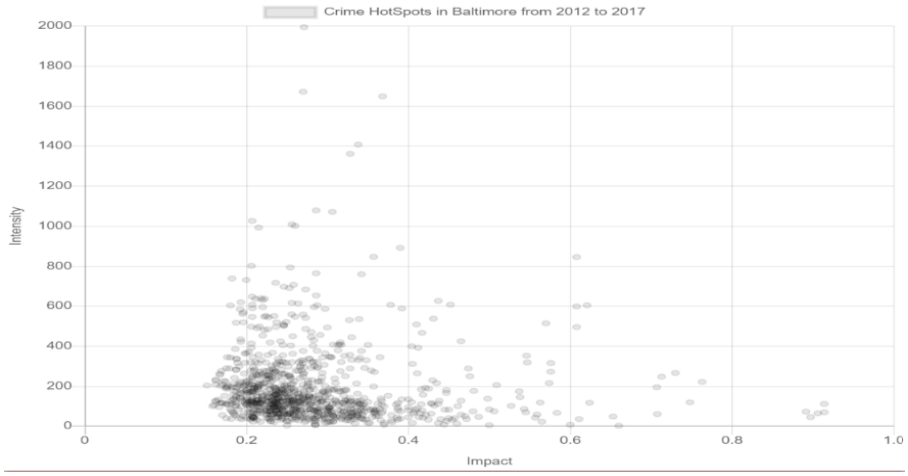


JavaScript d3:

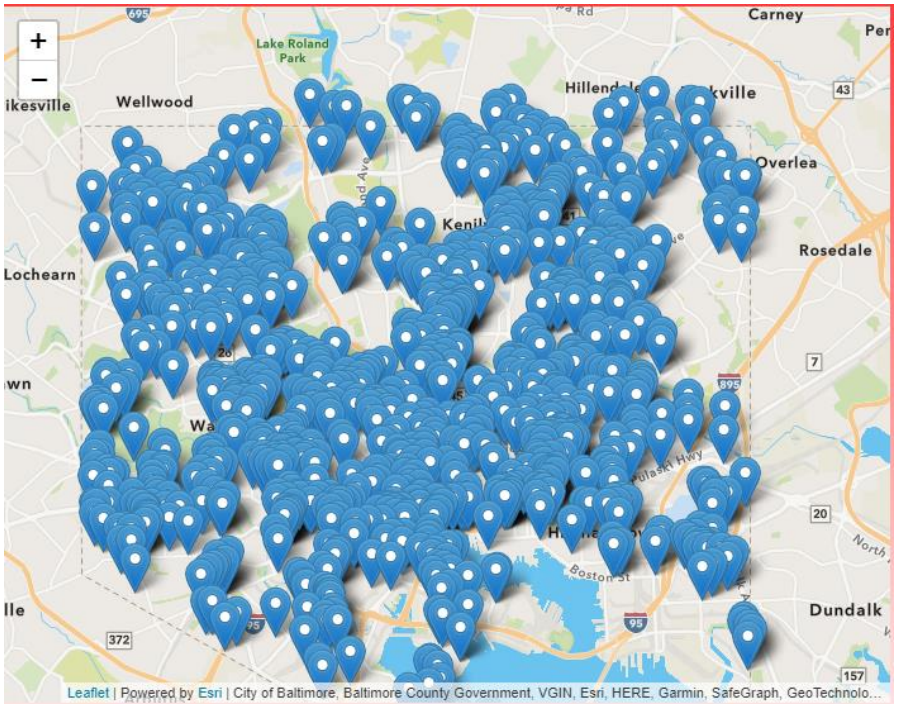
Initial visualization:



Final Visualization:

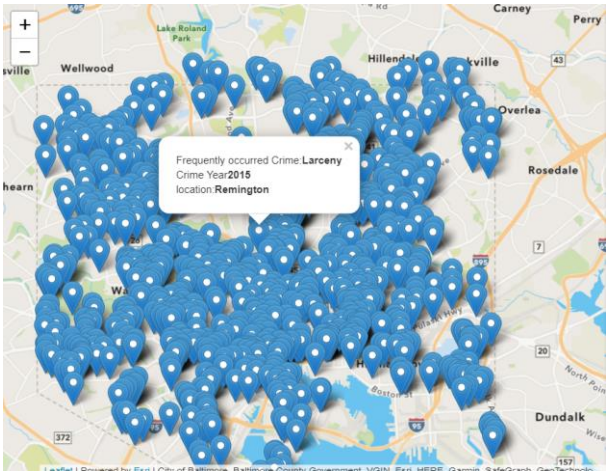


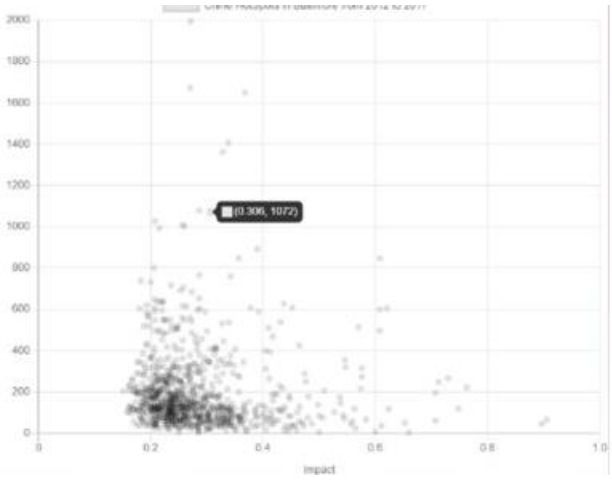
Given that our project is based on crimes in Baltimore City, showing the hotspots on the map would be fantastic because it would allow the data to be shown in a way that everyone could see and comprehend. We are now concentrating on how to use the leaflet library. First, we attempted to draw a map of Baltimore. We entered our dataset into the ArcGIS website to read the data using latitudes and longitudes in order to get a map. Following that, we drew a street-level map of Baltimore. The latitudes and longitudes of each hotspot were used to plot the locations on the map. We highlighted the crimes that had been committed most frequently in the area.



We added tooltips for all our visualizations based on comments made about our prototype. Dr. Federico recommended that we construct an interaction between the visuals. The implementation of a Scatterplot and Leaflet interface has begun. First, to generate dynamic interactions, we deployed a few JavaScript packages. For the dynamic interaction, the same map was employed. Then, using chart.js—which is also used for JavaScript visualizations—we select a scatter plot with a similar probability vs. intensity distribution. We constructed the street level of Baltimore City using an API link and Leaflet, and then we wrote a scatter plot code in the Leaflet code. We carried out dynamic interaction between the map and the scatter plot using various routines. We also preserved the possibility of being able to see those specific spots on the map if we zoom in closely. Additionally, we retained impact and

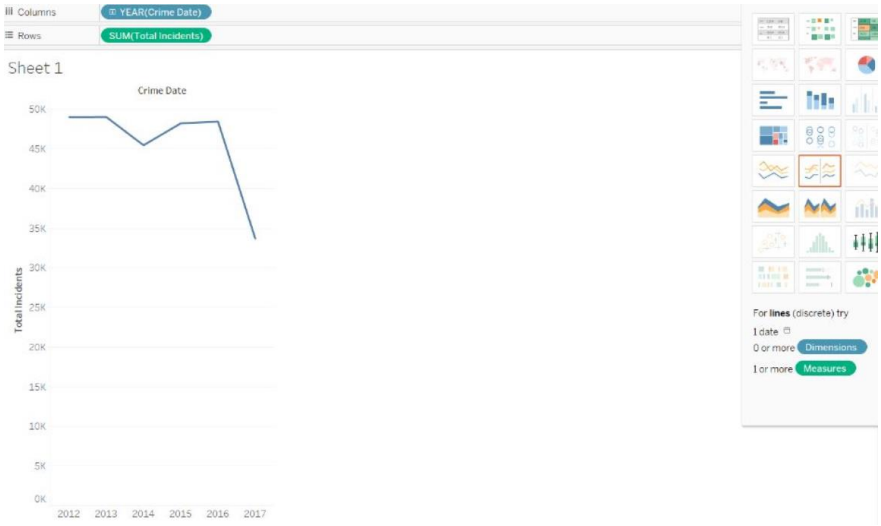
intensity on the scatter plot's tooltip. When we set the pointer there, we can see the effect and intensity of that location.

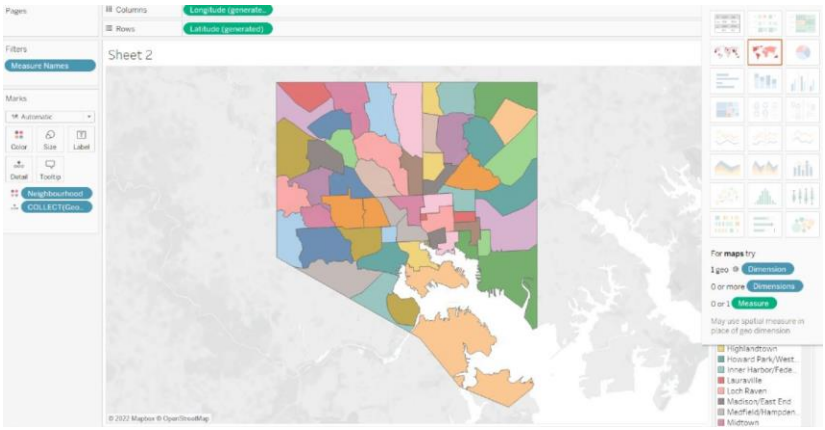


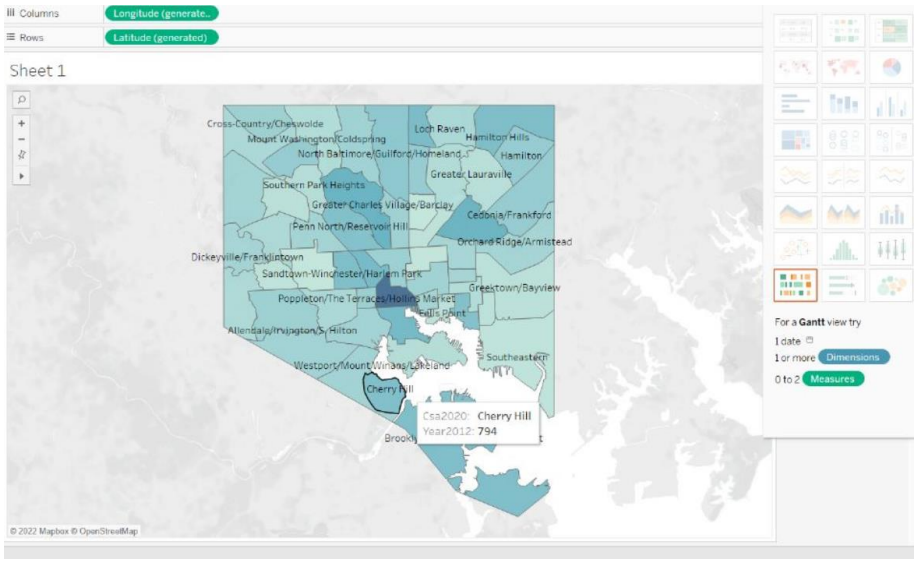


We have summed the severity of all crimes that have occurred in a neighborhood for our final interaction. We used the d3.geo equirectangular () method to produce the map of Baltimore. Baltimore City's boundaries were drawn using the community statistical Areas CSAs Reference Boundaries geojson file. The map was then made using some d3 functions. We used color hue and provided the range to illustrate the crime ranges in various areas. Then, to show the overall trends of the crimes in every area of Baltimore, we created the code for the line chart. We've decided to use the years as the X-axis and the crime count as the Y-axis. Now that the data has been added to the map, we can write the code for the line chart.

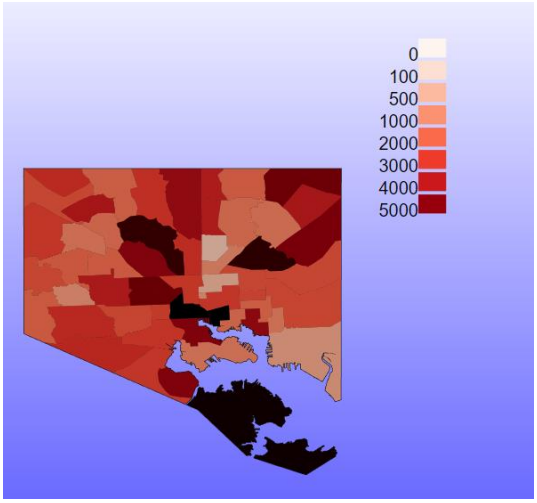
Tableau:

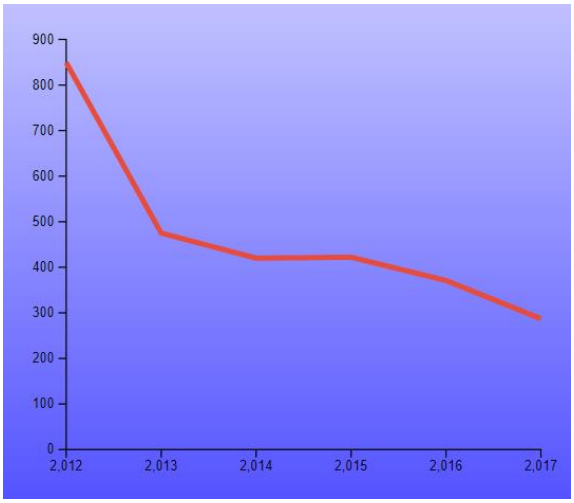

Columns
YEAR(Crime Date)
Rows
SUM(Total Incidents)
Sheet 1
Crime Date
Total Incidents
50K
45K
40K
35K
30K
25K
20K
15K
10K
5K
0K
2012
2013
2014
2015
2016
2017
For lines (discrete) try
1 date
0 or more
Dimensions
1 or more
Measures


Pages
Columns
Longitude (generate...
Rows
Latitude (generated)
Filters
Measure Names
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Neighbourhood
COLLECT(Geo...
Sheet 2
For maps try
1 geo
Dimension
0 or more
Dimensions
0 or 1
Measure
May use spatial measure in place of geo dimension
Highlandtown
Howard Park/West...
Inner Harbor/Fede...
Laursville
Loch Raven
Madison/East End
Medfield/Hampden
Midtown



Javascript D3:

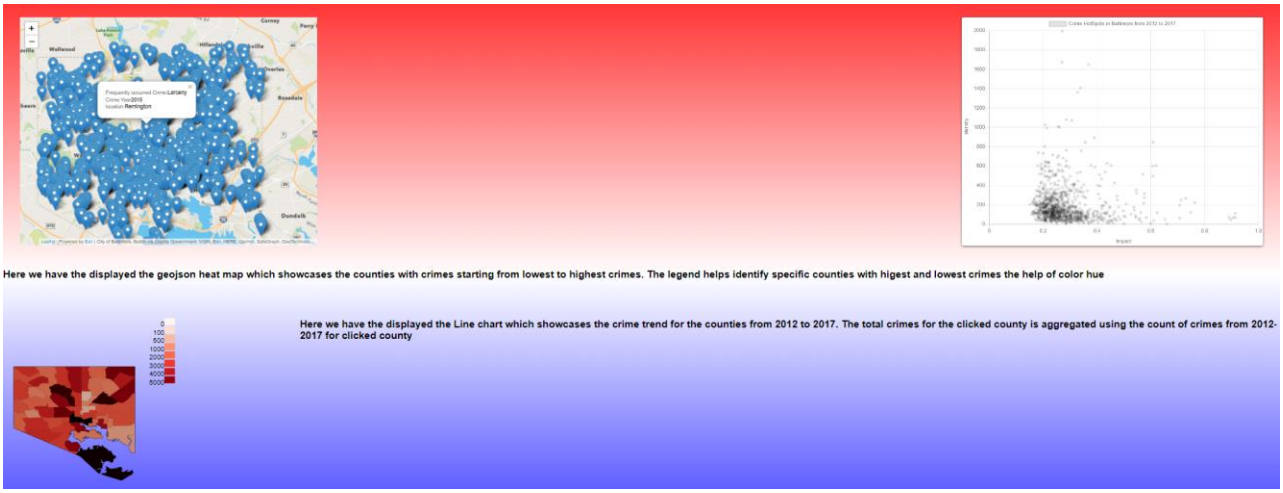


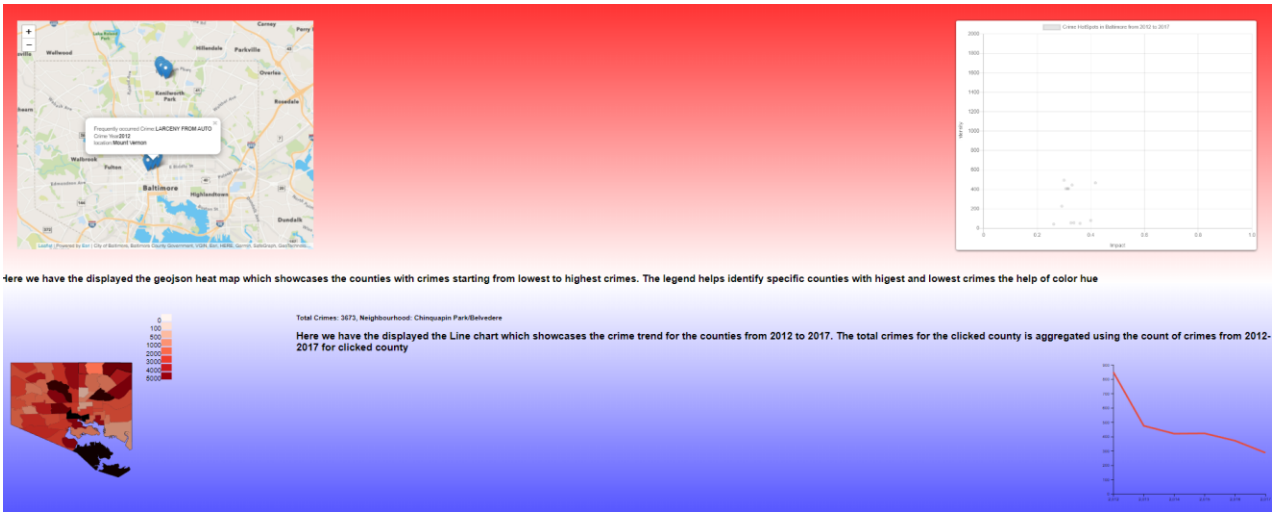


We were in the pipeline and made one adjustment as the professor had requested until our project was finished. Dr. Federico suggested that interactions be developed for all 4 visualizations simultaneously and we successfully created the interactions between all the visualizations. We have used scatter plot, two maps and then the line chart. (Scatter plot => Map & Map => Line chart).

7 Implementation:

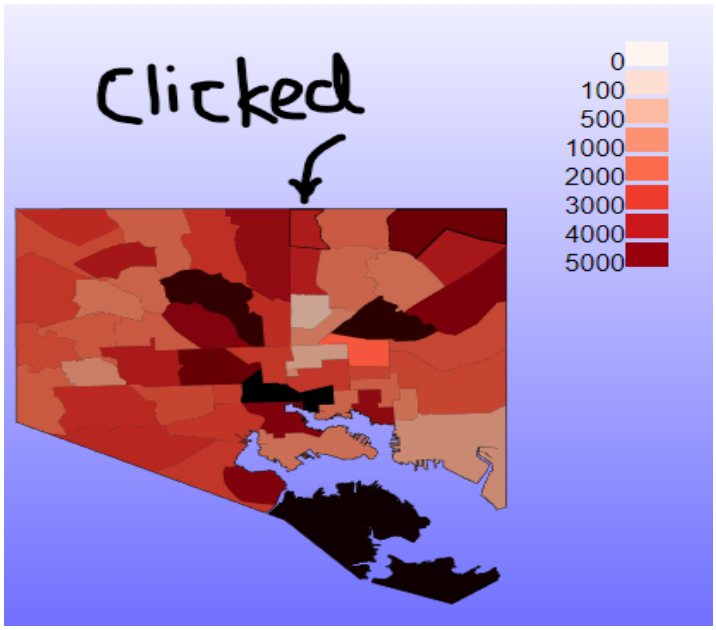
- We used the same leaflet-created map for the main visualization and selected a comparable probability vs. intensity scatter plot. To display the intensity and probability, we retained the tool tip on the scatter plot.
- Additionally, we retained the choice to filter hotspots in the scatter plot if the map is zoomed.
- In addition to the above visualization, we have shown the geoJSON map which has the counties and will display a line chart showing the total crime for the clicked county and trends of the crime data from 2012-2017.

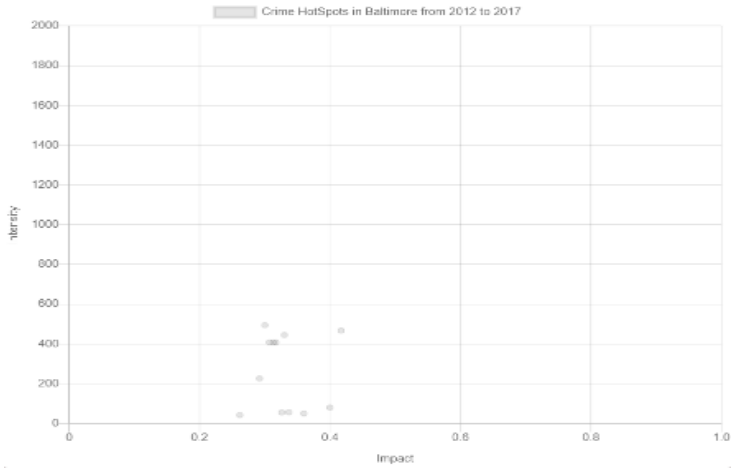


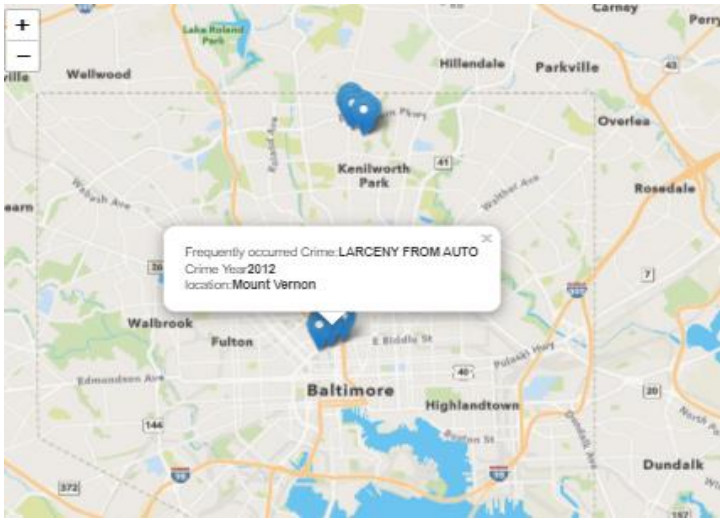


Here we have the displayed the geojson heat map which showcases the counties with crimes starting from lowest to highest crimes. The legend helps identify specific counties with higest and lowest crimes the help of color hue

Here we have the displayed the Line chart which showcases the crime trend for the counties from 2012 to 2017. The total crimes for the clicked county is aggregated using the count of crimes from 2012-2017 for clicked county

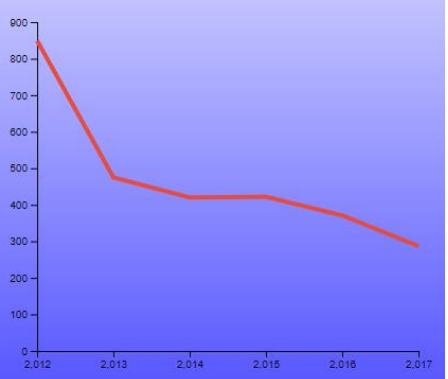






Total Crimes: 3673, Neighbourhood: Chinquapin Park/Belvedere

Here we have the displayed the Line chart which showcases the crime trend for the counties from 2012 to 2017. The total crimes for the clicked county is aggregated using the count of crimes from 2012-2017 for clicked county



- The geoJSON map is the main visualization which interacts with all the other visualizations using on click function.
- Once the click has been done the pop-up for neighborhood in that county will displayed on the leaflet, at the same time the scatterplot will be changed showing the intensity and probability of the pop-ups shown on the leaflet. Also, the line chart which shows the total crime and trends for crimes will also be displayed with the total crime and the name of neighborhood in that county as text besides the line chart.

8 Evaluation

Many insights are gained by carrying out these visualizations.

- First off, **"identifying the hotspots not on the total count of the crime but, on the probability of the crime that is likely to occur"** is the project's key selling point. The Baltimore city hotspots have been located by us.
- Second, we can see the multiple types of crimes committed in Baltimore.
- We can also see the precise spot where the crime took place. Let's assume that the incident happened somewhere like a grocery shop, street, etc.
- Finally, we can observe the overall patterns of the crimes over time.

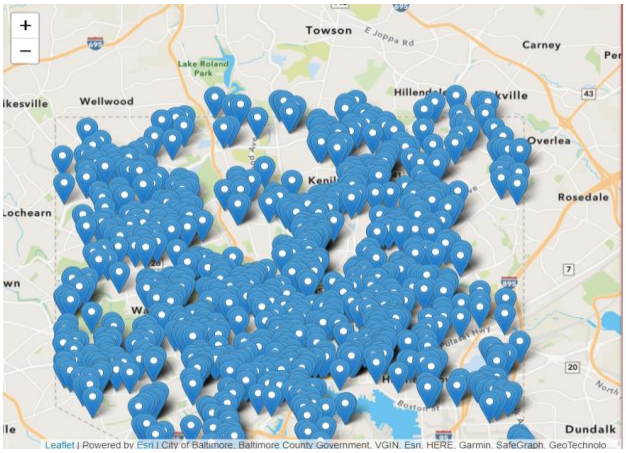
8.1 Working and Future Improvements:

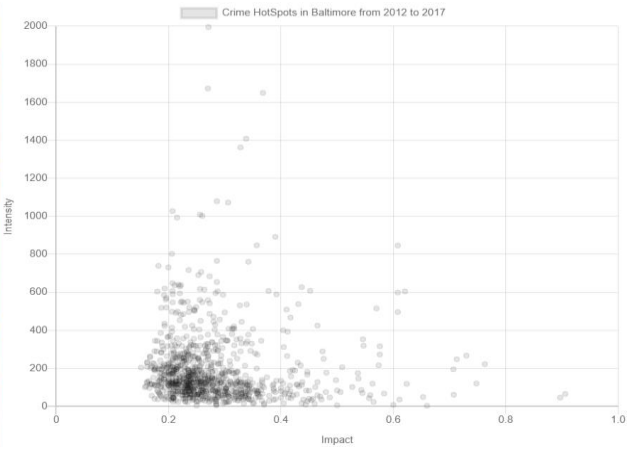
I. Using our visualizations, we can get the crimes data including location, total count etc. just by clicking on the counties which will show all the neighborhood in that county and shows the probability and intensity for the same location, also it is showing the total crime and the past crime trends.
II. This saves the time to go through the verbose documentation which consume time to provide the conclusion for any such crime data.
III. Furthermore, the same project can be expanded to other states/counties and help police department and other user to get the crime data instantly.
IV. If given a chance to work on the expanding the project, we can embed some python functions and deep learning techniques and models by showing the clusters and projection of hotspots and display the images which will enhance the project
V. To make the project scalable is our main goal.

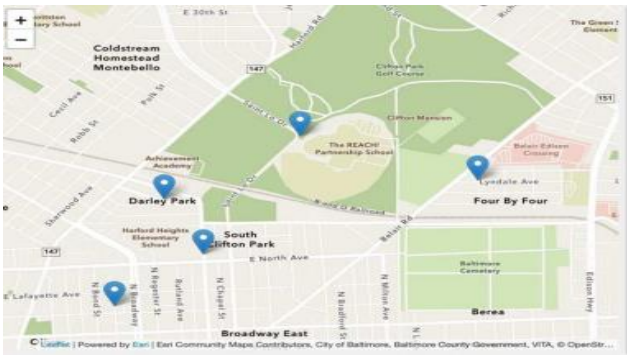
Our visualization work well by providing all information using few clicks and saves

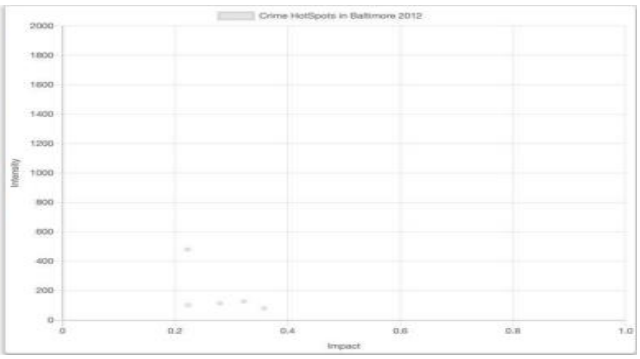
<u>Overall, our visualizations answered all the questions mentioned above</u>

- **Q1. What are the hotspots in Baltimore city based on the probability of the crimes?**

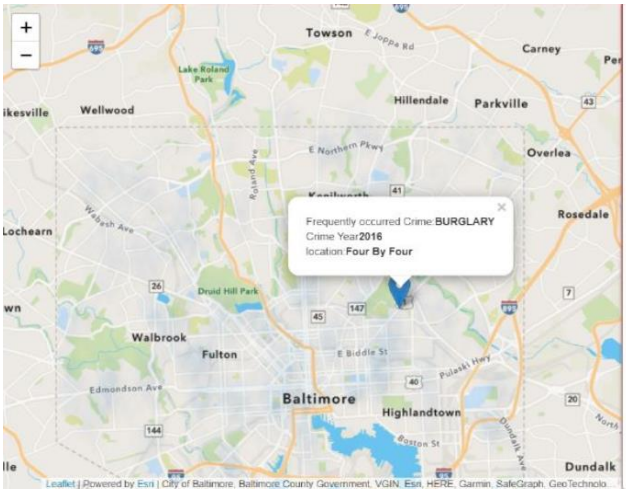


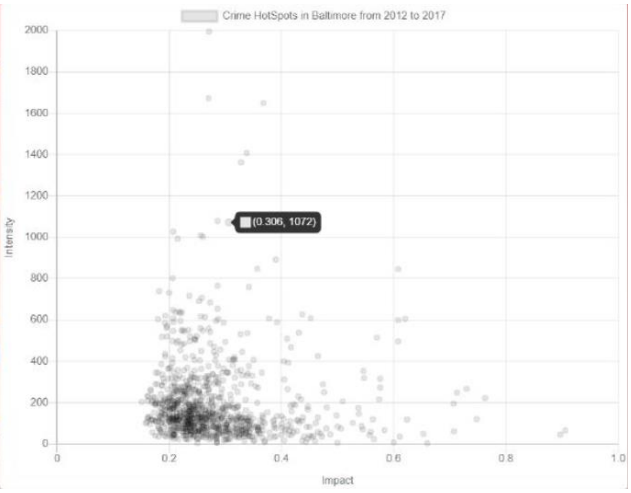




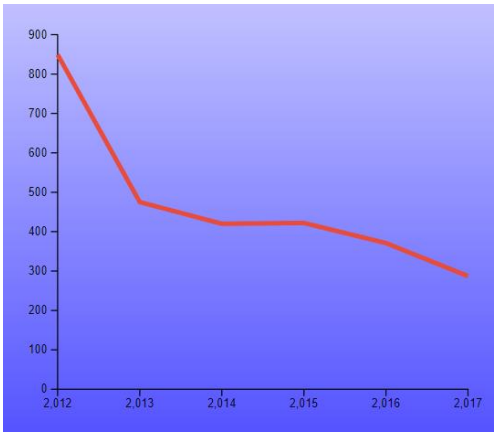


- **Q2. What is the crime that has the highest probability and intensity of occurring in a neighborhood?**

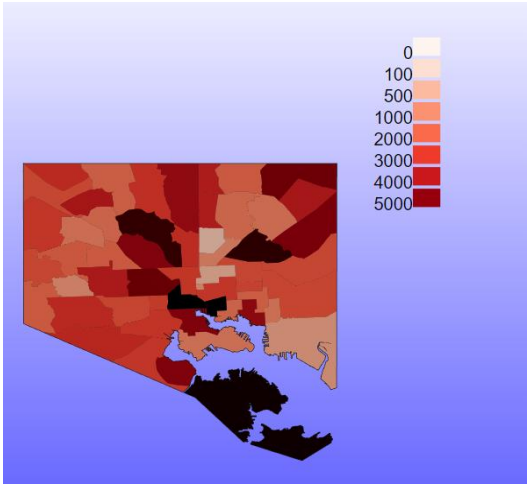


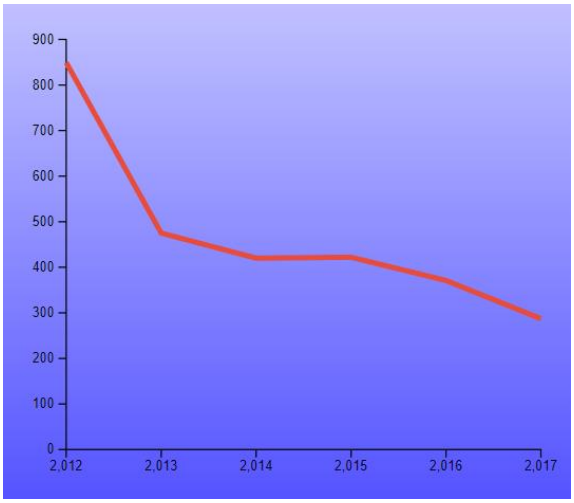


- **Q3. What is the overall trend of crimes over the years?**



- **Q4. Which area/neighborhood with total crime and trend from 2012-2017 crimes?**





By generating a scatter plot using the data from our first visualization, we were able to answer the first question and identify the hotspots in Baltimore. The dependent variable that we employed was the location's intensity.

The street-level map was created to help us find the answer to our second query. We've marked the place on a map and highlighted the crimes that have happened most frequently. We can clearly understand where exactly the crime has occurred and which crime has the most intensity by incorporating interactivity and tooltips between the street level map and the scatter plot.

By analyzing the trends throughout all Baltimore neighborhoods, we were able to respond to the final query by using the Line chart. Implementing the interaction makes the trends so obvious that we can clearly understand how they change over time.

By this point, we are confident that we have effectively completed the project and successfully responded to and implemented the process. We'd want to suggest that users can readily understand our data and will engage in the interactions we've designed.