US Crime Data Visualization

The link to our Github is here.
The link to our visualization is here.

Motivation and Goals

Our client, a family, is interested in relocating to either Los Angeles, California or Atlanta, Georgia for work. They want to know which parts of Los Angeles and Atlanta they should avoid moving to because of the high crime rate. They are also curious to know which city would be safer to live in as a family. We have decided to help our client by providing them with the visualizations from this project. If our client can identify the neighborhoods to avoid in each city and the safest of the two cities, then our data visualizations have served their purpose.

LA Visualization

In order to help our client make their decision for whether to move to LA county, we have implemented a linked geospatial leaflet, a stacked bar, and time series line plot in an easy-to-use Shiny application, which you can find here. Our data for LA ranges from 2010-2017 and to increase runtime we randomly sampled 10,000 observations. In Figure 1, the moveable panel on the left allows the user to enter up to three neighborhoods at a time and filter by the victim's age using the slider. The use of the slider can help our clients identify whether the amount of crime towards their age group in the selected neighborhoods is high. Based on the user's input, this will generate both a stacked bar plot which was created using plotly and a time series line plot in the same movable panel. Every time the user hovers over the stacked bar plot the proportion of crime in each of the three neighborhoods will show up where each color represents the neighborhoods. For instance when the user selects 77th Street, Devonshire and Wilshire as neighborhoods, selects ages between 21 and 98, and hovers over the "pink region" in the stacked bar plot for 2010, they will get information on the year, crime count , and neighborhood name, as seen in Figure 1. We can see that 77th Street has a crime count of 69 which is the highest number of crime in comparison to the other two neighborhoods. The user can easily depict which section of the bar has the highest crime count with their eye but the hover effect allows the user to get an exact number for that year. The line plot, on the other hand, shows the user the overall trend throughout the years and makes it easy to see the rise and fall of crime. Using the same example, the user can see that 77th Street had an overall high crime rate throughout all the years displayed. By observing just the bar plot and line plot, the user already knows that out of the three neighborhoods 77th Street wouldn't be ideal to live in because of its high crime rate. The movable panel gives our client an overall picture of what to expect but the leaflet map that is formed on the right allows the user to get more information and pinpoint these locations.

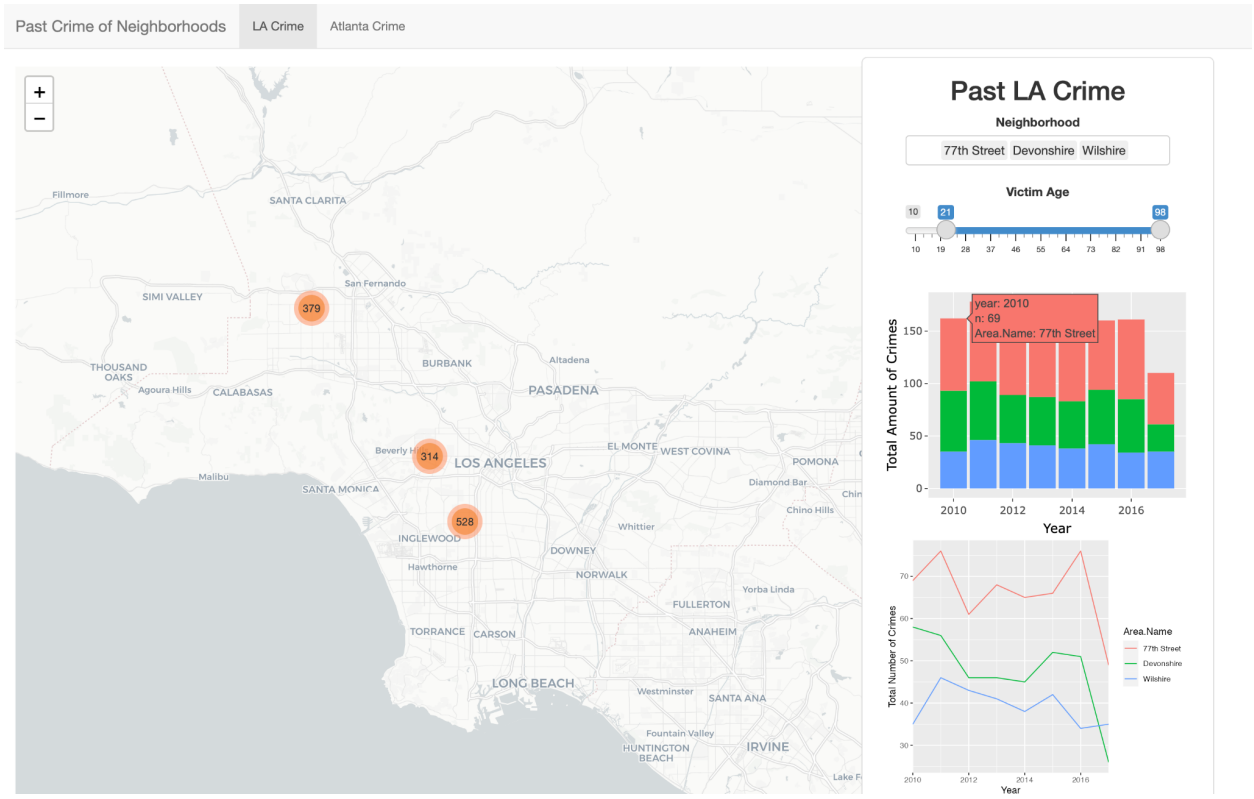

Figure 1: Movable Panel

Upon user input, the leaflet map will populate 'marker clusters' where each circle represents a neighborhood and the number on the circle represents the total crime count. Based on the information from the plots in the movable panel, the user knows that the circle with 528 as a total crime count represents 77th Street. As the user clicks on the circle, it will zoom in and update the amount of crime in that cluster until the points are no longer in a cluster. Figure 2 shows this transition between the clusters to single points.

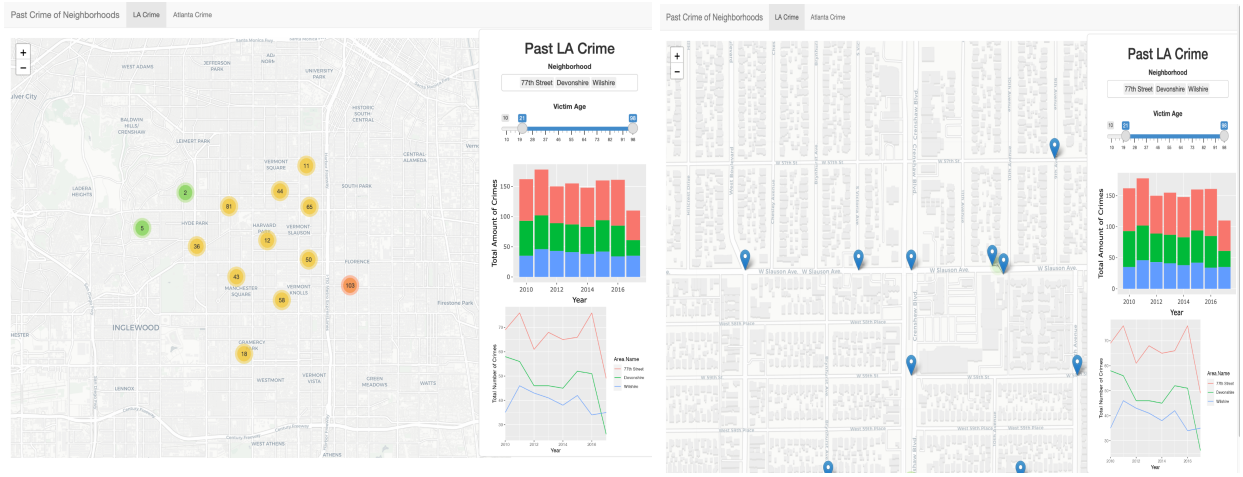

Figure 2: Transition from cluster to single points

The single points on the graph indicate the street where the crimes occurred. In the event where the family does decide to live in the 77th Street neighborhood, one of the streets they can avoid is W Slauson Ave. The user can also hover over each of the points to get information on the date the crime occurred, type of crime and victim's ethnicity, seen in Figure 3. All of the victims, from the data used, on W Slauson Ave were either Black(B) or Hispanic(H) residents who were assaulted or robbed. The information displayed to our clients when they hover over can allow them to determine whether their ethnic group is a target. It will also remind them to be careful with their interactions and to avoid possessing valuable items in places with high assault and robbery.

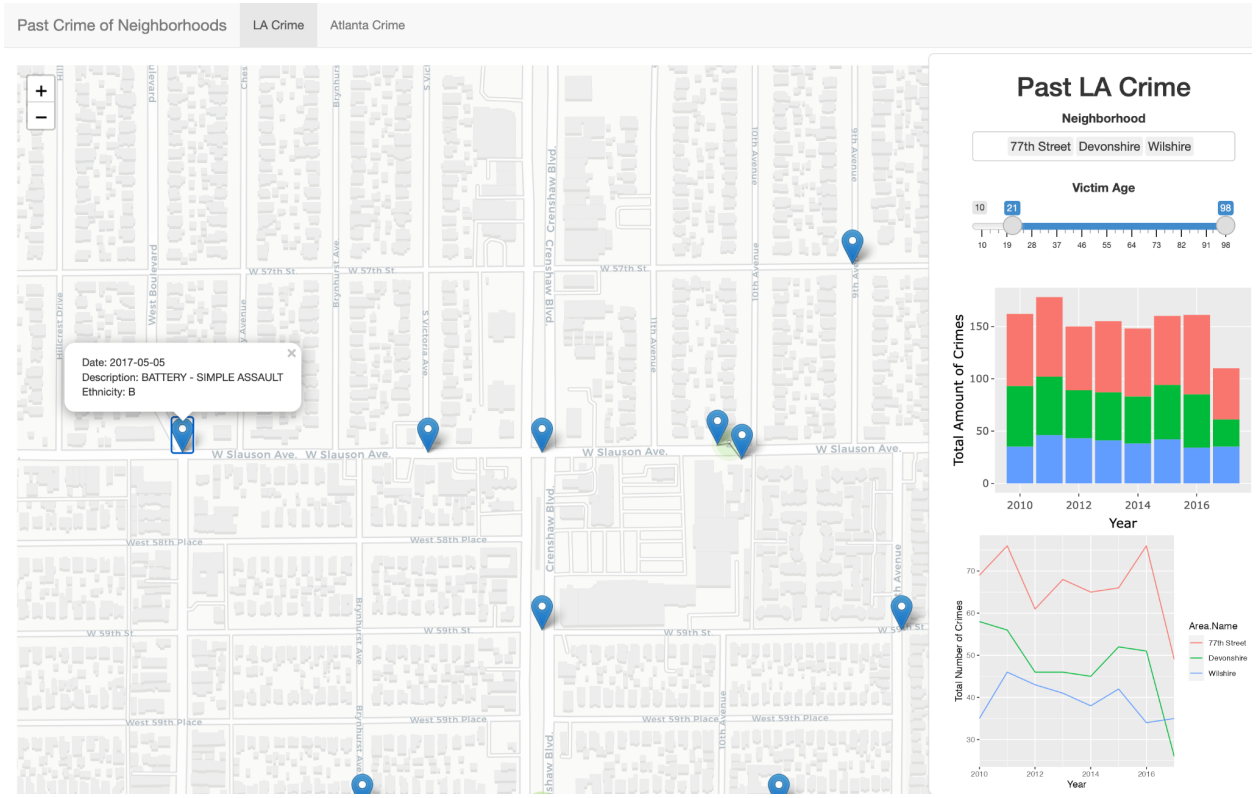

Figure 3: Information generated by a single point

Overall, our visualization will give the client information on which neighborhoods and streets to avoid. It will also give them information on the victim's ethnicity and age to help them eliminate themselves as possible targets.

Our prototype from Project Milestone 2 had multiple limitations, such as: little information displayed, the proportions of the two visualizations did not look nice, and the markers were very cluttered especially when more than one neighborhood was selected. Figure 4 demonstrates this. We corrected these issues to make it easiest for our clients to read our visualizations accurately. The one trade off is that since we limited the number of neighborhoods the user can select to

three, it will take more time to go through each set of neighborhoods. Our group, however, feels that this trade-off is minor and the overall cleanliness of the final visualization is worth it.

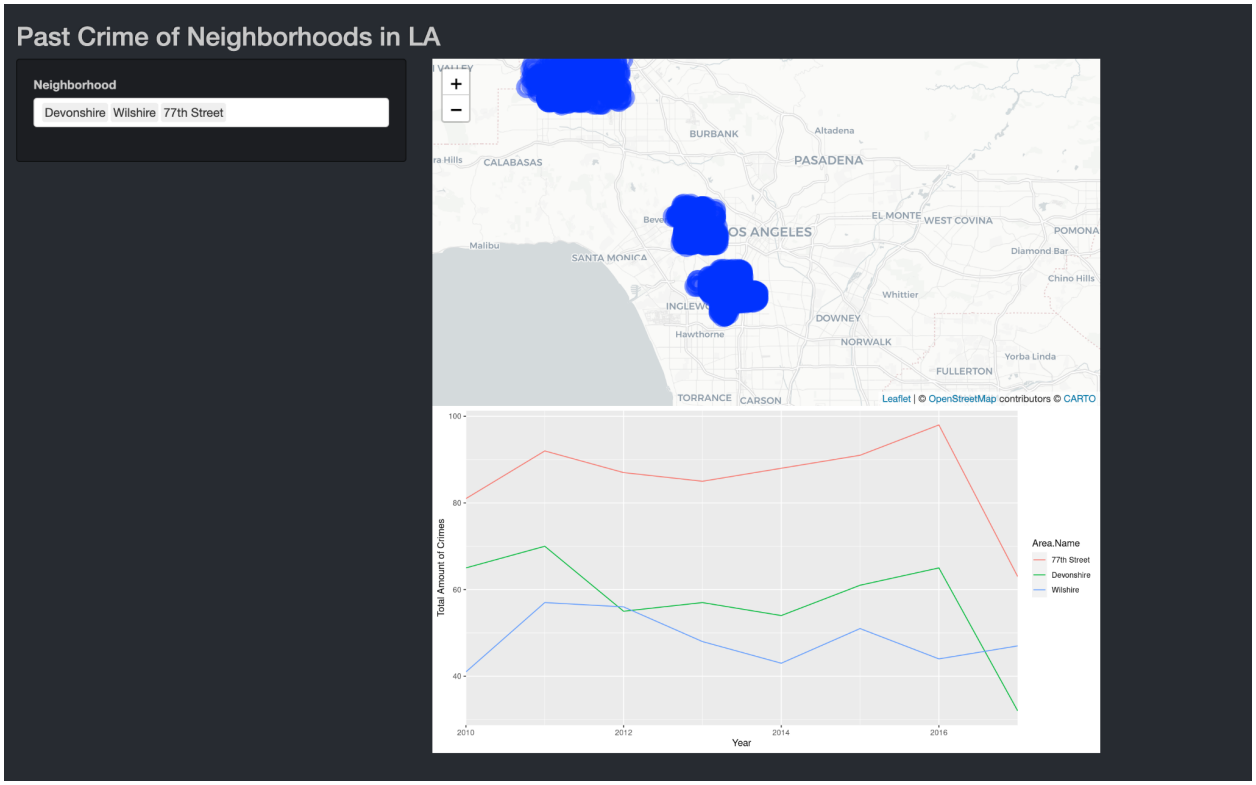

Figure 4: Prototype

Throughout this project we have explored many options for displaying past crime in order to achieve our goal and help our client. Our main sources of inspiration stemmed from Andrew Wheeler and George Usuwu. As mentioned in Project Milestone 2 we used Wheeler's advice of using a long term scale instead of short term so the client could see the overall trend. He also mentioned that plotting the categories of the variables is more useful than plotting together. We implemented this idea by allowing the client to select different neighborhoods instead of looking at them as a whole. Additionally, we took Usuwu's advice regarding density maps to display past crime. Instead of a heatmap, however, we utilized the clustering method for our markers on the geospatial map. Each circle on the map displays the density, or total number of crime, in that cluster, so our client would want to stay away from the clusters with the highest number displayed in the circles.

Atlanta Visualization

The Atlanta map differs from Los Angeles. Because of limitations in the data set, we decided to focus on displaying crime patterns. We took inspiration from Makin and Bye's analysis of sex crime and sex industry activities (Makin, 2016) and created a geospatial visualization that

attempted to highlight the areas where certain types of crime is more likely to be committed. This places an emphasis on viewing the overall patterns of crime versus the specific information of each individual crime that the Los Angeles map features. We decided that color is the most appropriate tool of distinguishing crime types from each other and allowing the user to most obviously see the areas where certain crimes are more rampant.

Figure 5 shows our final implementation of this geospatial interactive map. First, the user needs to select an Neighborhood Planning Unit (NPU). The City of Atlanta lists on their website the twenty-five different urban planning districts that all neighborhoods in this city are categorized into. Then, one will be able to select a number of neighborhoods to compare. To view the crime data, the last selection is the type of crime the user is concerned with. Instead of labels, we found it easier in our final implementation to create a legend indicating which color represents which type of crime (11 total types found in this data set). In this example below, our NPU X has a small neighborhood called Capital View Manor and a relatively larger one further south known as Hammond Park. We can see that pedestrian robbery is frequently seen in Hammond Park, but there have only been a few cases of pedestrian robbery on the outskirts of Capital View Manor between the years 2009 and 2016. On the bottom, our line graph indicates the changes in the total amount of crime in each neighborhood throughout the same years (2009-2016). One can conclude that overall crime is much higher in Hammond Park than in Capital View Manor, but it is also necessary to use the map visualization to realize the size of these neighborhoods are not equal, and so we may not be able to make a definitive assumption that one area is necessarily safer than the other. Also, crime rates in both neighborhoods were decreasing between 2015 and 2016, and if both trends continued in the more recent years then both places could be good options for our client to consider.

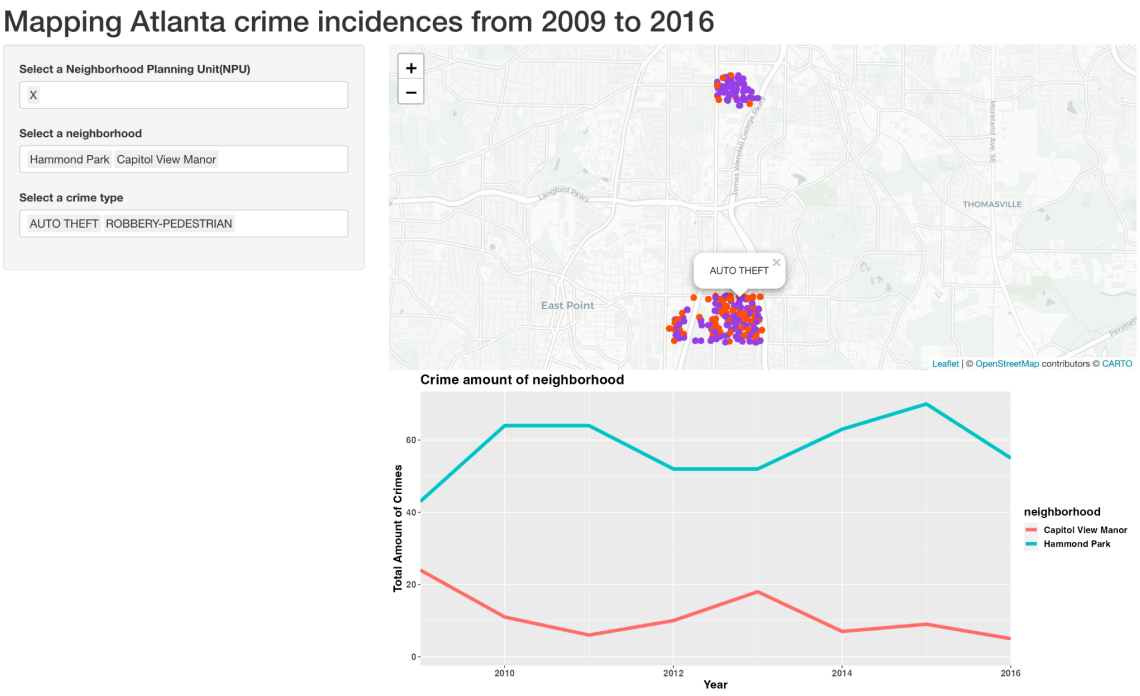

Figure 5: Atlanta Visualization

The main challenge of our Atlanta visualization is the difficulty in being able to easily zoom in and out of the different selected neighborhoods. There is no automatic zooming because we chose to not cluster our data in the same way Los Angeles is clustered in the previous section, since this will make it difficult to color each point the way we would like for each crime to be displayed. The line graph, while informative of crime frequency over time, also has some drawbacks. It was mentioned before that the total number of crimes may be extremely dependent on the area of a neighborhood district, and we did not incorporate any standardization to take neighborhood size into consideration. It is also difficult to view the same types of crimes that are shown on the map in the line graph, as we are losing information when summing these frequencies to a single value per year.

Conclusion

To conclude, we would like for our client and users of the application to realize that the visualizations for these two cities seem similar, but the map features provide very different information. Both have their pros and cons, and it may be best for users to view both cities using both types of visualization (although we would need more specific information on the nature of crimes committed in Atlanta that are given in the LA labels). Future steps would be to find convenient ways of zooming in on Atlanta neighborhoods and NPUs, additional D3 or Shiny implementations of interactive line graphs, incorporating time series data into our geospatial visualizations, and looking more methods of visualization for the the data of victims of different

types of crime. We may also generalize these visualizations to other large cities, depending on the client's interests.

References

Bryant, A. (2017, March 18). *Crime in Atlanta 2009-2017* . data.world. Retrieved October 7, 2022, from https://data.world/bryantahb/crime-in-atlanta-2009-2017

City of Los Angeles. (2017, September 18). *Crime in Los Angeles*. Kaggle. Retrieved October 7, 2022, from https://www.kaggle.com/datasets/cityofLA/crime-in-los-angeles

David A. Makin & Caroline Bye (2018) Commodification of Flesh: Data Visualization Techniques and Interest in the Licit Sex Industry, Deviant Behavior, 39:1, 46-63, DOI: 10.1080/01639625.2016.1260383

Ratcliffe, J.(n.d). *Aoristic analysis*. Jerry Ratcliffe. Retrieved October 8, 2022 , from https://www.jratcliffe.net/aoristic-analysis

Sankaran, K. (n.d.). *Geospatial Interaction* . Statistical Data Visualization. Retrieved November 18, 2022, from https://krisrs1128.github.io/stat679_notes/2022/06/01/week8-4.html

Sankaran, K. (n.d.). *IMDB Shiny Application* . Statistical Data Visualization. Retrieved November 18, 2022, from https://krisrs1128.github.io/stat679_notes/2022/06/01/week2-3.html

Usuwu, G. (n.d.). Crime Geography. Retrieved October 14, 2022, from https://www.researchgate.net/publication/343721767_Crime_Geography

Wheeler, A. P.(n.d). Crime Data Visualization for the Future. https://www.dropbox.com/s/hphfswijruwjz8q/DataViz_Future_AndyWheeler.pdf?dl=0