Seattle Crime History Analysis and Distribution

HCDE 511 Info Visualization, Autumn 2018, Final Submission by Team C.H.A.D.
Alexander Van Roijen, avanroi1@uw.edu Cynthia Feng, cynfeng@uw.edu Mengxiao Song, mxsong@uw.edu Prakhar Agarwal, pa2511@uw.edu

LINK: *Hosted Visualization*

1. EXECUTIVE SUMMARY

Team C.H.A.D. sought to provide an improved interactive visualization experience to better understand the publicly available Seattle Police Department crime report dataset. [1]. We combined data on population from the U.S. American Community Survey [2] to allow us to understand how crime rates may change along with an increasing Seattle population. We ultimately created a Tableau tool that allows users to explore crime per capita as well as percentages of particular crime subcategories on the granularity of Seattle neighborhoods or SPD beats with further filtering metrics on time of day and years of interest. We intend these tools to be useful for policy makers and curious citizens. We end this submission with a few examples of our exploration of the data to understand how policies and laws such as the legalization of recreational marijuana may impact DUIs and narcotic cases throughout Seattle.

2. INTRODUCTION

As Seattle residents, our team was interested in exploring data related to crime rate and type within Seattle over time. We wanted to understand different crime rates and types in different neighborhoods. After much discussion we decided we did not want to bring in demographic data and social issues into our visualization, because that was outside the scope of what we were trying to achieve, and risked overstating our data by adding our own bias and opinions. Our goal from the start was to explore the Seattle Police Department (SPD) records of official police reports filed in their database. Our potential audiences were the public for visualizing crime rate and type in different neighborhoods of Seattle, as well as the police department themselves for visualizing where they could heighten patrol at different times of day, for example.
The SPD's official website had publicly available datasets, which included a record of all initial police reports taken by Seattle police officers. This dataset, called Crime Data, contains offense categorization coded to simulate the standard categories reported to the FBI under the National Incident Based Reporting System (NIBRS). It contains police reports for occurrences as early as 1908, with the data being most complete starting 2008, when the department began logging all reports in the Departments Records Managing System (RMS). This is the same continuously updated dataset that is used in SeaStat (https://www.seattle.gov/police/information-and-data/seastat) meetings to address crime statistics in Seattle.

The dataset, extracted mid-November 2018, had about 500,000 unique police reports organized into rows. The columns were report number (for lookup in the police report viewing system), occurred date, occurred time, reported date, reported time, crime subcategory, primary offense description, precinct, sector, beat (boundaries for patrol deployment), and neighborhood [1]. The SPD also offered spatial data files containing .shp files (See Figure 1) organized by police beats that we could upload into Tableau to create mappings. The crime dataset was too massive, resulting in slow processing time when we uploaded it into Tableau. We did some data processing to clean up the dataset for faster dynamic queries for our final visualization.

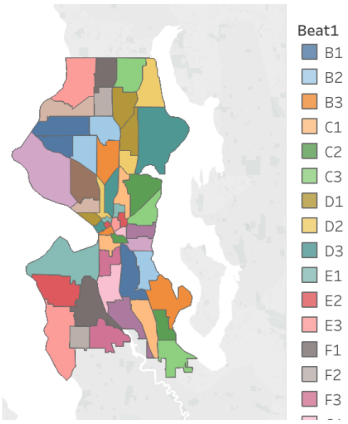

Figure 1. Mapping created from SPD's spatial file of police beats (patrolling boundaries)

The crime dataset was too massive, resulting in slow processing time when we uploaded it into Tableau. We had to do some data processing to clean up the dataset for faster dynamic queries for our final visualization.

3. Related Works

1. **SPD Crime Dashboard**.SPD has released their own interactive Tableau visualization. Notice that you can toggle with any parameter in the original dataset except for the time of day the crimes may have occurred. This is likely due to the size of the dataset with 450,000 rows slowing down responsiveness. We seeked to put this feature in our visualization. The dashboard includes tooltips when hovering over the neighborhoods that detail the top five prevailing issues or concerns within that region. There is lastly a monthly crime chart below that shows how crime has changed based on the filters selected.

2. **Seattle Times Article**. Using the Seattle crime data set, the Seattle Times released an article highlighting the change in property and violent crime in the neighborhoods of Seattle between 2010-2012 and 2015-2017. The viewer starts with an overview of all crimes and can zoom into either violent crimes or property crimes and see how different neighborhoods fared over the five year period. Clicking the neighborhoods tell you the exact change in crime rates of different types and the neighborhood rank compared to the others.

4. Design Process

4.1. Problem Space

Safety is a big issue for everyone, and sometimes the real crime rate can not be correctly perceived by the public. In the article from Seattle Times, it was inspiring for our team to see the data showing Seattle crime rate has been declining in the recent years, which the public may not be aware of. One of our initial questions was, how might we help the public as well as policy makers to better understand the real crime situation?

4.2. Potential Audiences and Desired outcomes

Our potential audiences are policy makers, the Seattle Police Department, and people who live in Seattle. For policy makers, the tool may help them visualize the impact of policies on crime rate, and help them with future policy making. For current or future Seattle residents, the tool may provide an unbiased perspective for the public to visualize crime rate in different neighborhoods, and may help them with decision making, such as where to lease an apartment.

4.3. Data Cleaning

The first step was to include Seattle neighborhood population data. This proved quite difficult, as the estimates from the U.S. census only define populations to the granularity of Çensus Tracts", which are boundaries determined by the census to contain approximately 2000 persons each. Consequently, there is no nice overlap between census tracts and neighborhood boundaries that are defined somewhat haphazardly. Thus, the process of matching census tracts with their appropriate neighborhoods was challenging. Utilizing python pandas, the task was ultimately completed, but overestimated the population of most neighborhoods as census tracts were shared between multiple neighborhoods.
The second step was handling the temporal data. We binned the temporal data in chunks of 2 hrs each, so a day was divided into 12 chunks. See Table 2 for the binning results in the time_hr column, in which 0 corresponds to 12:00AM to 2:00AM, 2 corresponds to 2:00AM to 4:00AM, and 22 corresponds to 10:00PM to 11:59PM. We also shrunk the dataset from over 450,000 rows to 188,000 rows by grouping crime subcategories and occurred times and beats, and adding a c̈ountc̈olumn for number of reports of that category. This resulted in significantly faster processing time when uploading to Tableau.

4.3.1. DATA PROFILE

1. Original data (4̃50,000 rows x 11 columns)

   Nominal: Primary Offense Description, Neighborhood, Beat, Precinct, Sector

   Ordinal: Occurred Date, Occurred Time, Reported Date, Reported Time

   Quantitative: Population, Count

2. Processed data : Created a crime records dataset from data.csv file (1̃88,000 rows x 7 columns) . *This consists of reports of crime data from the year 2008 - present.*

   Nominal: Crime Subcategory, Neighborhood, Beat

   Ordinal: Year, time_hr

   Quantitative: Population, Count

4.4. Ideation Process

Our design process started with brainstorming. Each of the team members came up with conceptual ideas in a limited time activity. After discussion, we found that we wanted to focus on categories of crime, occurred year, occurred time of day, neighborhood, population, and crime trends in our final visualization.

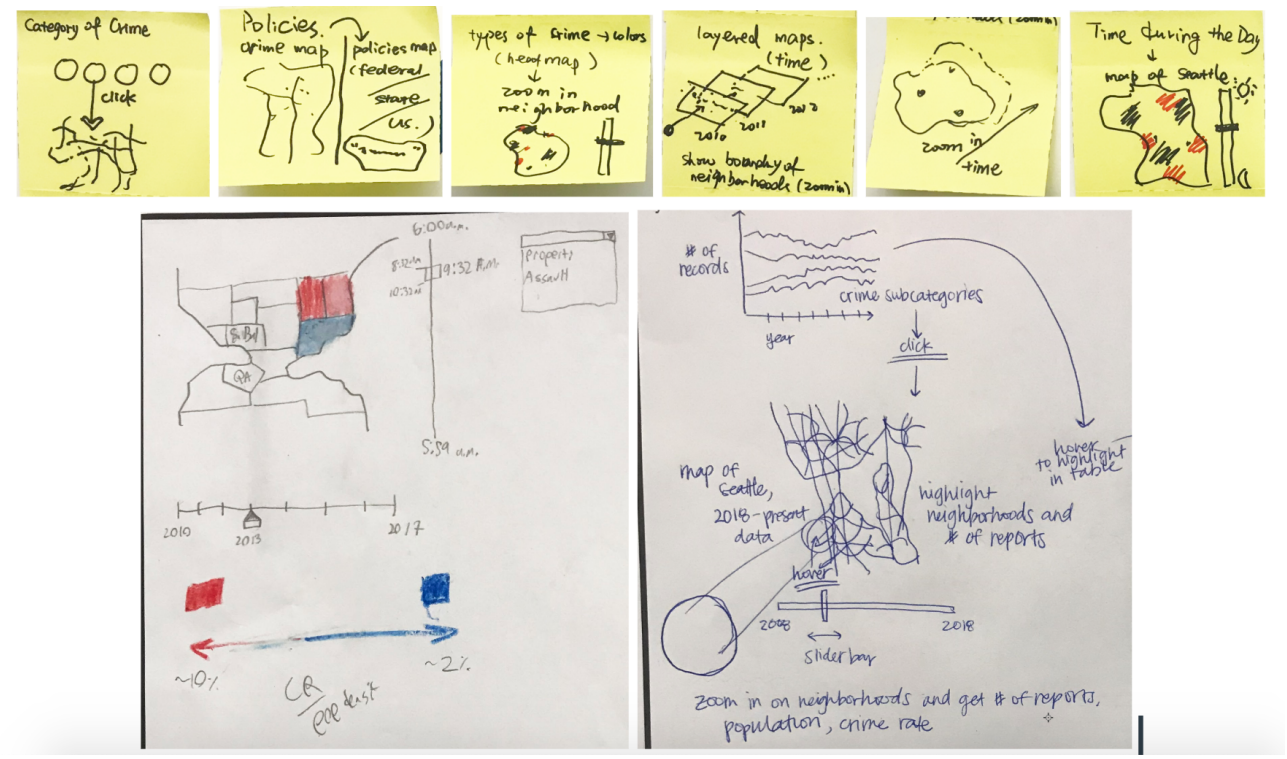

Figure 2. Initial ideas for final visualization

Then we created our low fidelity prototype on Tableau. It showed neighborhood, crime categories, occurred year, and occurred time during the day. We tested it with three participants (P1, P2, and P3) outside the class. We gave them tasks to go through the prototype and tried to get valuable feedback from them.

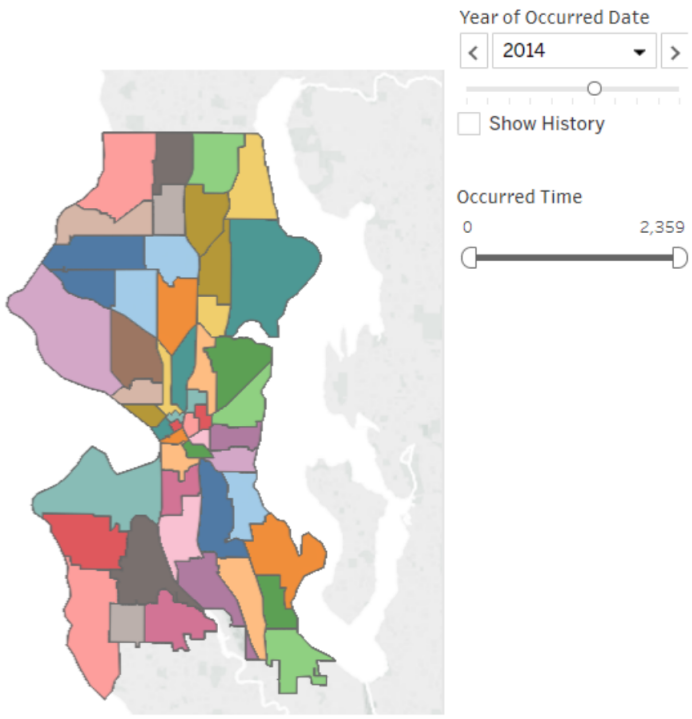

Figure 3. Low-fi prototype on Tableau

One of our key findings was the slow performance negatively impacted user experience. P1 was confused about the slider for filtering of time which read from 0 to 2,359. P3 was skeptical about the accuracy of trends of crimes, and she expressed a desire to have the ability to compare certain neighborhoods over certain years

to help her with decision making.

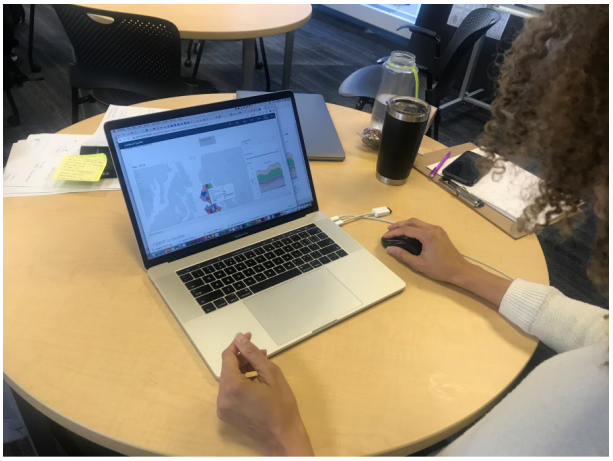

Figure 4. Participant testing our low-fi prototype

After analyzing all the feedback from testing, we refined our prototype based on our participants' insights and came up with the final design solution.

4.5. Final Design Solution

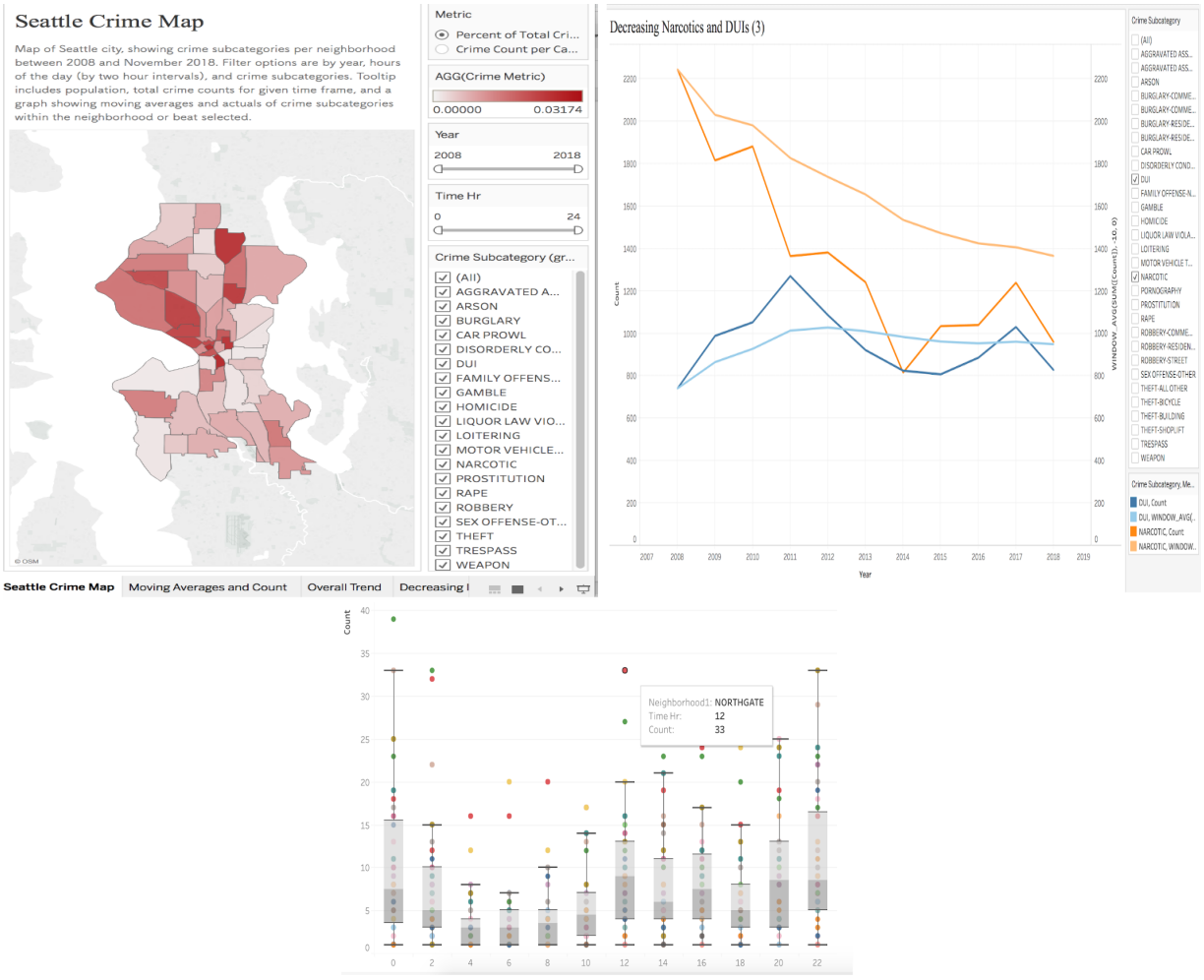

Figure 5. Final visualization

In our final design solution, users can see a heatmap of crime using filters to select neighborhoods, categories of crimes, occurred year, and occurred time during the day. By clicking a neighborhood on the map, the user could zoom into a line plot of the filtered crime subcategories, as well as moving averages of those crime subcategories over the years chosen. This is valuable insight to see the general trend in crime count. There is also an option to visualize the data in percent of total crime, or crime count per capita.
By using dynamic queries, this visualization tool allows users to rapidly, safely, and playfully explore a database[3]. Users are able to start with an overview mapping of the data, and at the same time, they can explore and specify known item queries related to occurred year, occurred time, and crime subcategory.
This visualization follows the Visual Information-Seeking Mantra. Users are able to have an overview of the data, and then zoom in by clicking. When hovering on different neighborhoods, more information will appear in the tooltip that provides a closer view. Then, based on this, users can use filters to select specific information they need.
Our final design solution also applies the Congruence Principle. The Congruence Principle suggests that the structure and content of the external representation should correspond to the desired structure and content of the internal representation.[4]. For example, using line plots to show the crime trend, and using a deeper color for a higher rate of crime. This principle helped make the visualization more understandable to users.

5. Results

Our work was to provide a visual tool for the crime trends in different neighborhoods of Seattle. We used an array of visualizations to cover varying granularity of details in our analysis.

1. HeatMap overlaid on Seattle's map: This provided a generalized overview of high crime areas based on the crime category chosen using the filters.

2. Box plot across bucketed time: This provides an overview of distribution of crime across varying hours during the day. Using the filters provided, we can further investigate the distribution for selected crimes categories in selected areas in Seattle.

3. Moving Averages on a line chart: This provides estimate of trends of crime over a period of time in selected areas and selected crime categories.

Using some of the base tools we further investigated four individual use cases in greater detail:

5.1. DUI and Narcotics

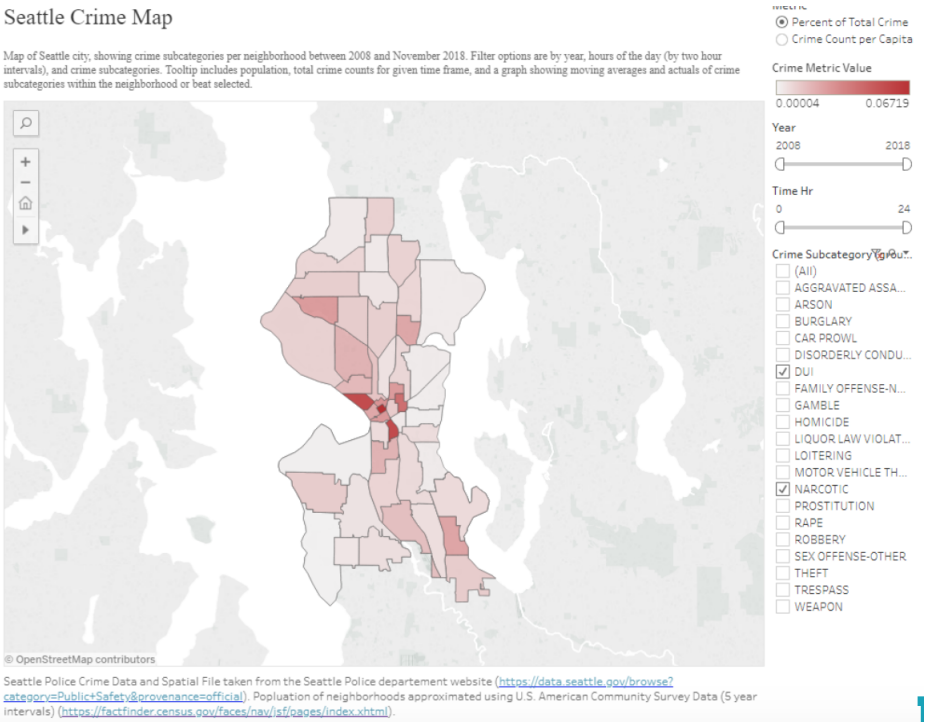

Figure 6. Percentage of DUI and Narcotic Rates in Seattle over a decade

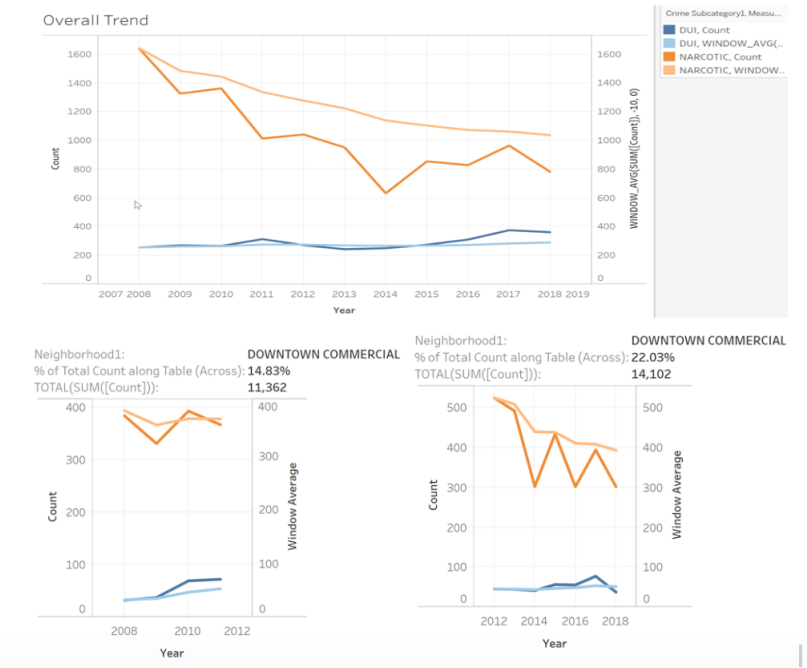

Figure 7. Seattle and Downtown Commercial shows a steady decreases in Narcotics. Particularly past 2012

In November of 2012, amendments were made to legalize the recreational sales of marijuana and consequent usage. Using the tool, we found results that correlated with suspicions that legalization of marijuana would reduce reports of narcotics. There was suspicion that DUIs may increase with relaxed laws on marijuana usage, however, these results seems to indicate that these counts have remained rather fixed.

5.2. Violent Crime

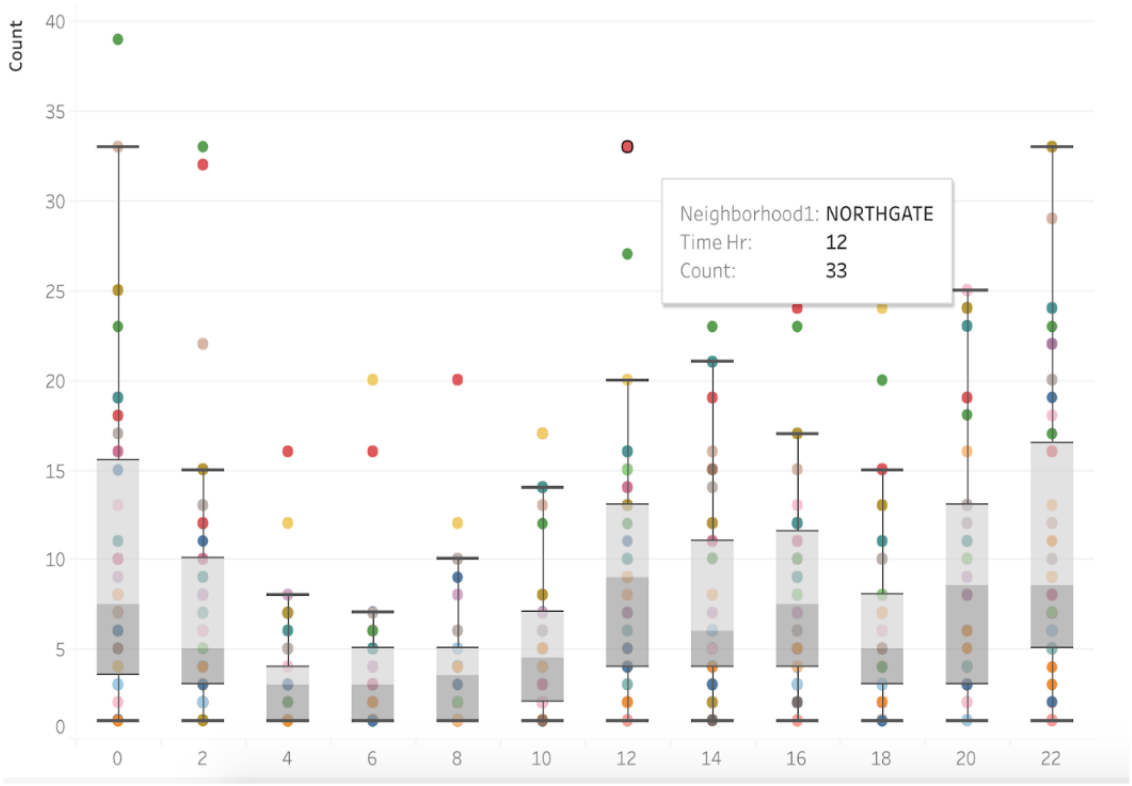

Figure 8. Violent Crime

As can be seen from the above plot, early hours happen to be the top choice for criminals. We also find that Northgate and Downtown neighborhoods are the hot spots for violent crimes.

5.3. Weapon Crime

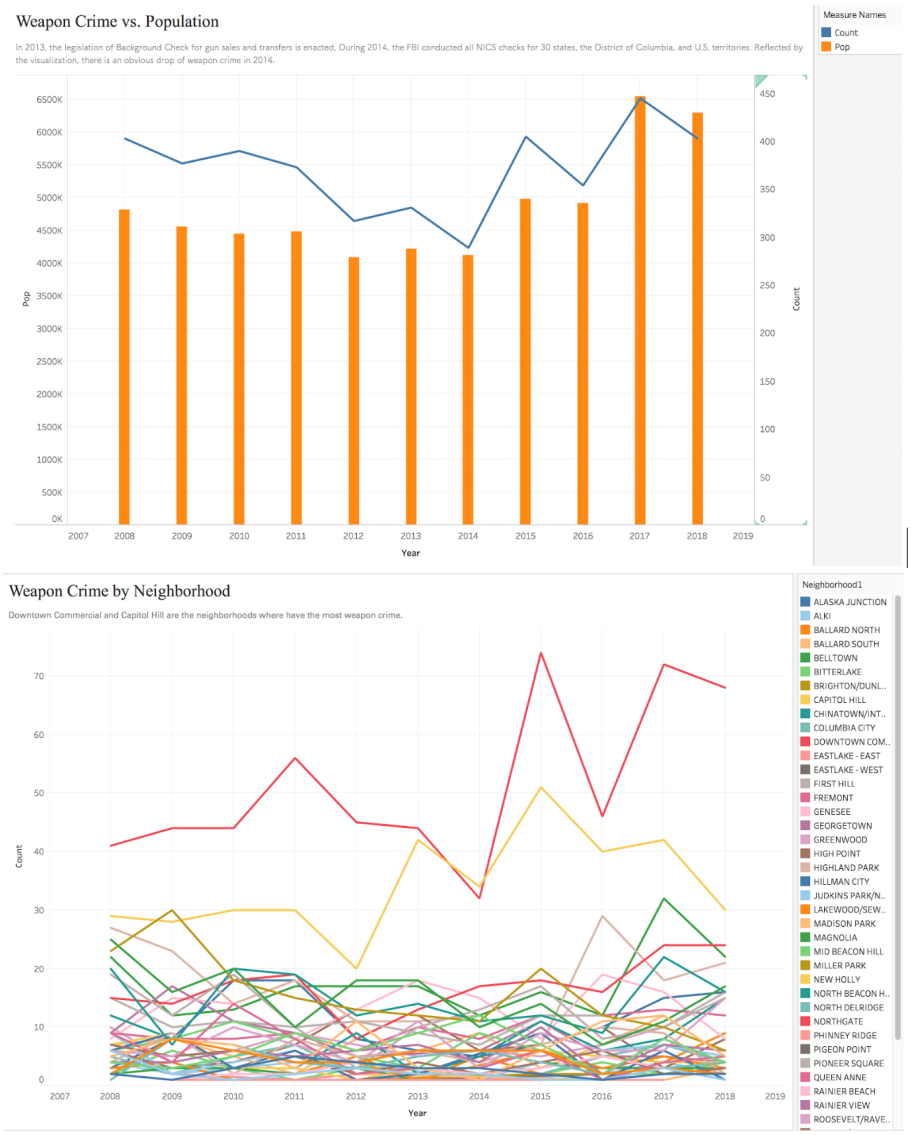

Figure 9. Weapon Crime with Population and Neighborhood

In 2013, the legislation of Background Check for gun sales and transfers was enacted, During 2014, the FBI conducted all NICS checks for 30 states, the District of Columbia, and U.S. territories. Reflected by the visualization, there is an obvious drop of weapon crime in 2014, while the population does not drop that much during that year. So the policy of background check for gun sales may contribute to the decrease of weapon crime. In this use case, we see that given a context of what has happened can help reduce the bias.

5.4. Property Crime

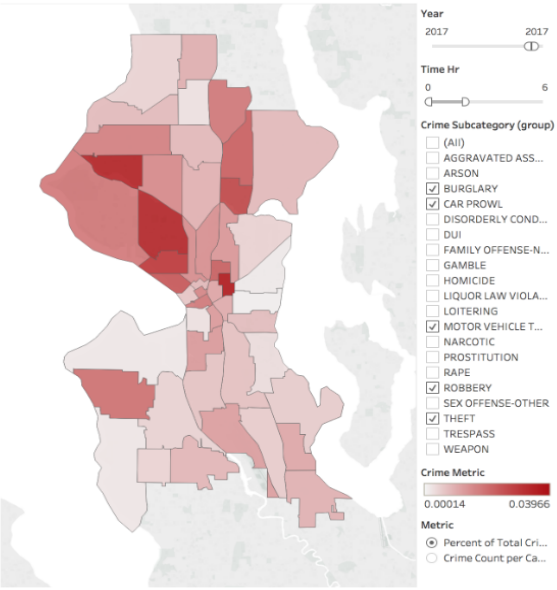

Figure 10. Property Crime heatmap

This was a deep dive into property crime, which consisted of burglary, car prowl, motor vehicle theft, robbery, and theft, between the hours of midnight and 6:00AM in 2017. The map showed that Capitol Hill had the most property crime as a percent of total crime, almost 4.0 %, with the Queen Anne neighborhood with second highest property crime rate.

6. Discussion / Future Work

6.1. Incorporating more population metrics

We want to include more data with regards to population in more visuals. Often times, we can see crime rates increase and assume its for the worst. However, if the population of that region were to increase significantly, it would sometimes make sense to see a similar increase in crime. Further, commercial areas that receive more traffic are more likely to see vehicular related incidents.

6.2. Transitions

We have some useful graphics on display, particularly when you click on the neighborhood of interest. However, it will likely confuse users as there is no smooth transition or labeling to clearly indicate what they are viewing in relation to their previous map of Seattle. There are also other labeling issues we seek to resolve in the near future.

7. REFERENCES

1. Seattle Police Department crime report dataset., 2018.
2. U.S. American Community Survey.
3. Card, M., *Readings in information visualization: using vision to think*; Morgan Kaufmann: 1999.
4. Tversky, B.; Morrison, J. B. y Betrancourt, M. *International journal of human-computer studies* **2002**, *57*, 247-262.

Appendices

Original Data Extract vs. Cleaned Data

Report Number	Occurred Date	Occurred Time	Reported Date	Reported Time	Crime Subcategory	Primary Offense Description	Precinct	Sector	Beat	Neighborhood
20080000465209	12/13/1908	2114	12/13/2008	2114	DUI	DUI-LIQUOR	EAST	G	G2	CENTRAL AREA/SQUIRE PARK
20080000445053	11/26/2008	1445	11/26/2008	1445	THEFT-SHOPLIFT	THEFT-SHOPLIFT	NORTH	U	U2	UNIVERSITY
20080000443239	11/26/2008	1440	11/26/2008	1440	THEFT-BUILDING	THEFT-BUILDING	EAST	E	E1	CAPITOL HILL
20080000443265	11/26/2008	1412	11/26/2008	1412	TRESPASS	TRESPASS	WEST	K	K3	CHINATOWN/INTERNATIONAL DISTRICT
20080000443330	11/26/2008	1410	11/26/2008	2038	CAR PROWL	THEFT-CARPROWL	NORTH	N	N1	BITTERLAKE
20080000467804	11/26/2008	1408	12/16/2008	1408	BURGLARY-RESIDENTIAL	BURGLARY-FORCE-RES	WEST	D	D2	SLU/CASCADE
20080000482505	11/26/2008	1400	12/31/2008	905	THEFT-ALL OTHER	THEFT-OTH	WEST	M	M2	DOWNTOWN COMMERCIAL
20080000461992	11/26/2008	1400	12/11/2008	1115	THEFT-ALL OTHER	THEFT-OTH	NORTH	U	U3	ROOSEVELT/RAVENNA

Table A1. Sample of original Crime Data set

year	Crime Subcategory	Neighborhood	population	Beat	time_hr	count
2008	AGGRAVATED ASSAULT	ALASKA JUNCTION	17421	W2	0	2
2008	AGGRAVATED ASSAULT	ALASKA JUNCTION	17421	W2	10	1
2008	AGGRAVATED ASSAULT	ALASKA JUNCTION	17421	W2	12	1
2008	AGGRAVATED ASSAULT	ALASKA JUNCTION	17421	W2	14	3
2008	AGGRAVATED ASSAULT	ALASKA JUNCTION	17421	W2	2	2
2008	AGGRAVATED ASSAULT	ALASKA JUNCTION	17421	W2	20	1
2008	AGGRAVATED ASSAULT	ALASKA JUNCTION	17421	W2	22	3
2008	AGGRAVATED ASSAULT	ALASKA JUNCTION	17421	W2	6	1

Table A2. Sample of data after cleaning and binning for faster processing time, with population data added