



2009-2019

# Seattle Crime

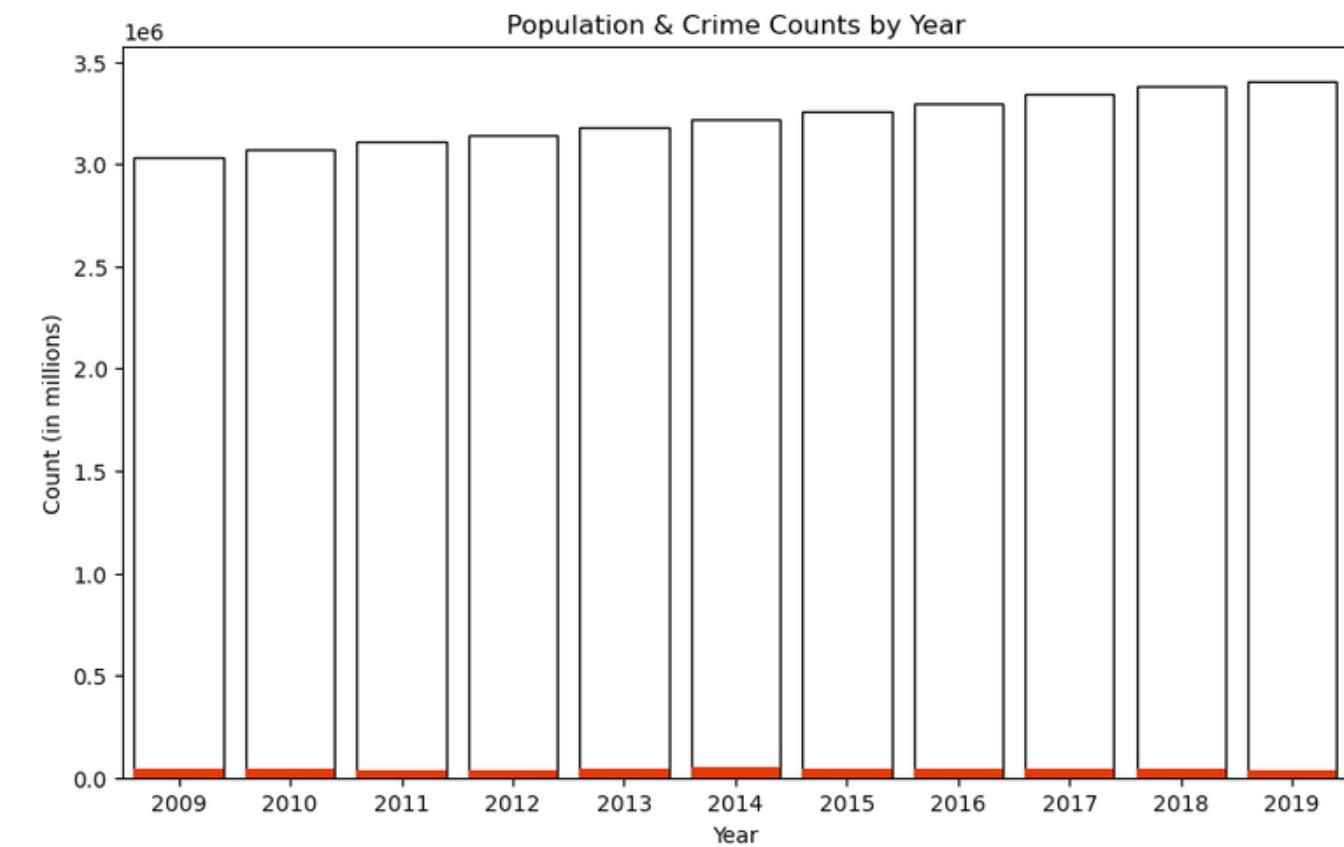
An analysis by Ana K. Santiago Monroy

# Executive Summary

*Main questions: Is the city really as dangerous as it is portrayed? What type of crimes are most predominant?*

## Insights:

- Relative to Seattle's population, one's probability of experiencing **any** type of crime in **all** of Seattle ranged between **1.1 - 1.5%** from 2009-2019.
- The primary type of crime reported across all districts is 'Property' crime.
- The district that saw the most number of offenses across this decade was Downtown.
- The district with the least number of offenses was Seward Park.



The white bars represent Seattle's population, while the crime counts for each year are depicted in red.

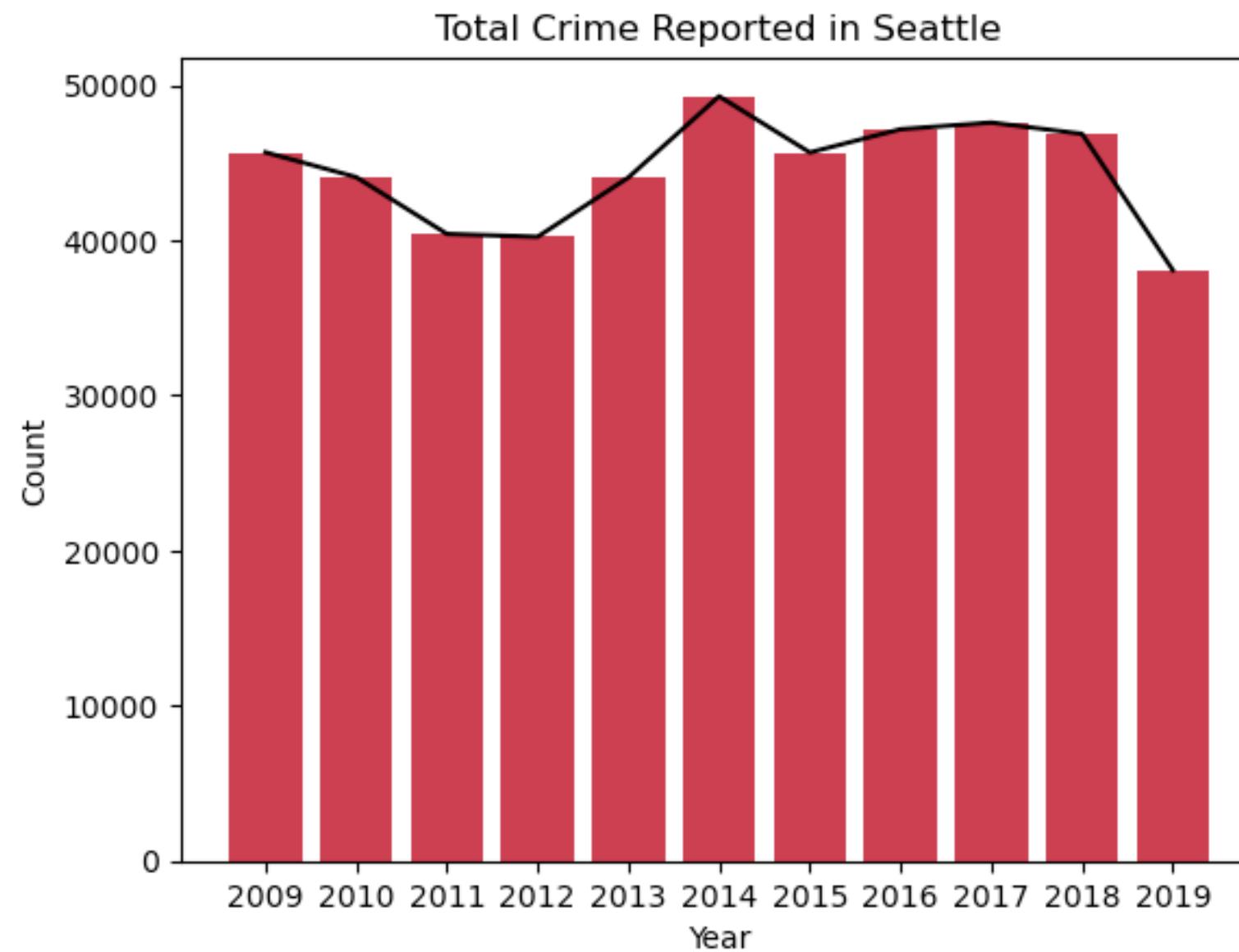
# Methodology

- Examined Seattle crime data from 2009-2019
- Utilized **API keys** made available by data.seattle.org
- Joined data from secondary sources (Seattle GeoData & MacroTrends.net)
- Used the following **libraries**: Pandas, Requests, Matplotlib, Seaborn, JSON, Math, Numpy, Scikit-Learn, FuzzyWuzzy
- Ingested and cleaned Seattle crime data, joined this source with district information to narrow down areas of focus for analysis.
- Conducted EDA to define pertinent questions.
- Visualized insights in Python, utilizing Seaborn and Matplotlib

Jupyter notebook hyperlink



# EDA



My first couple of questions for this analysis were:

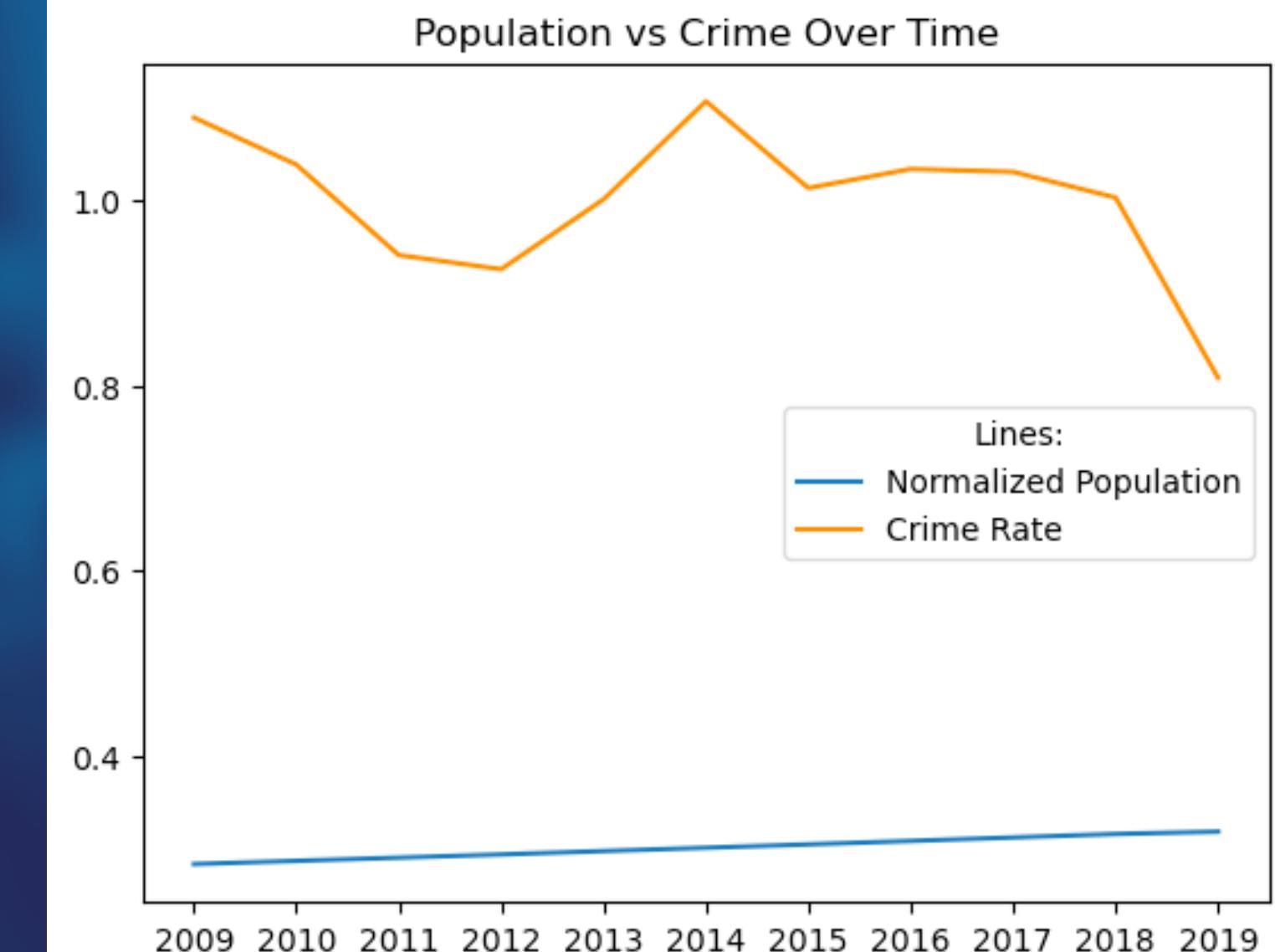
- 1) Does crime in the city followed a linear trend?
- 2) What the total amount of crimes in a given year?

This visualization reveals that the assumptions from these questions are not true. Firstly, and surprisingly, the trend line began to decrease between 2018 and 2019.

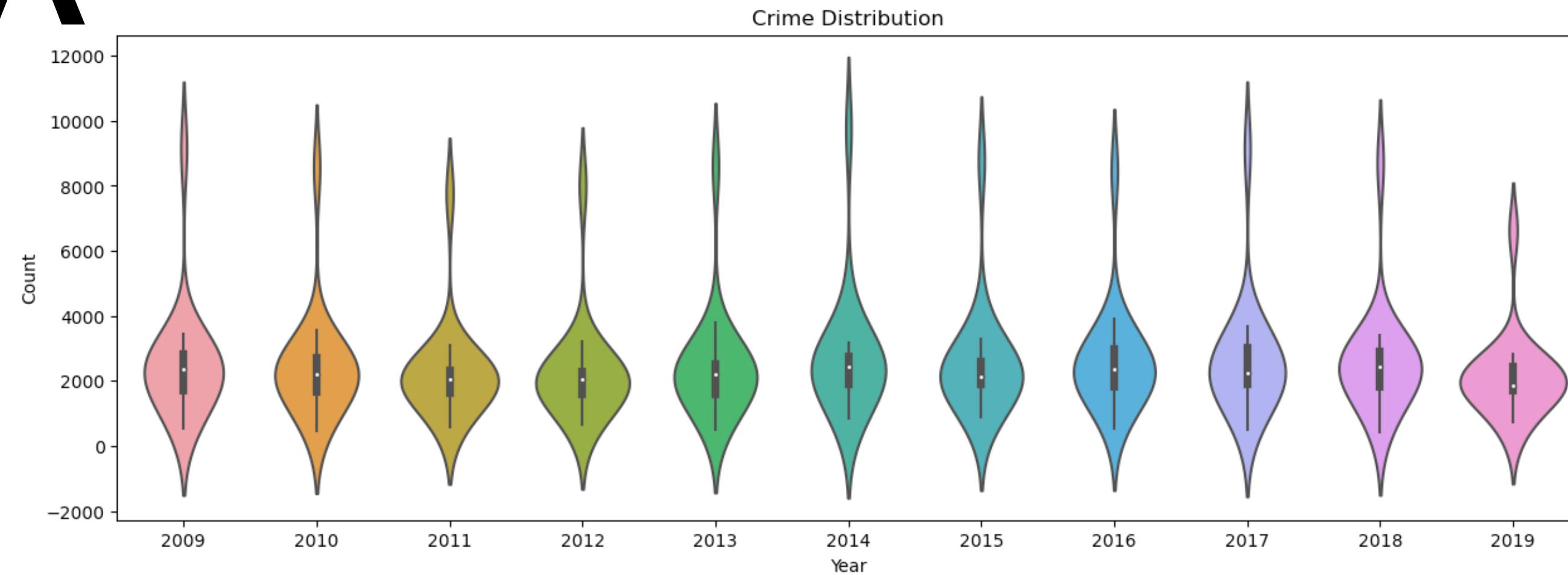
Additionally, the maximum number of crimes reported in a single year across all of Seattle was less than 49,000.

# EDA

When comparing population and crime count values, the resulting visualization proved challenging to interpret, especially when assessing any potential relationship between the two variables. After standardizing these values, I opted to plot them to observe the shape of their respective trend lines and ascertain whether population increases influence the crime rate. The plot reveals relatively linear population increases, yet notably, the crime rate remains unaffected by these changes, exhibiting its own distinct pattern marked by significant fluctuations.

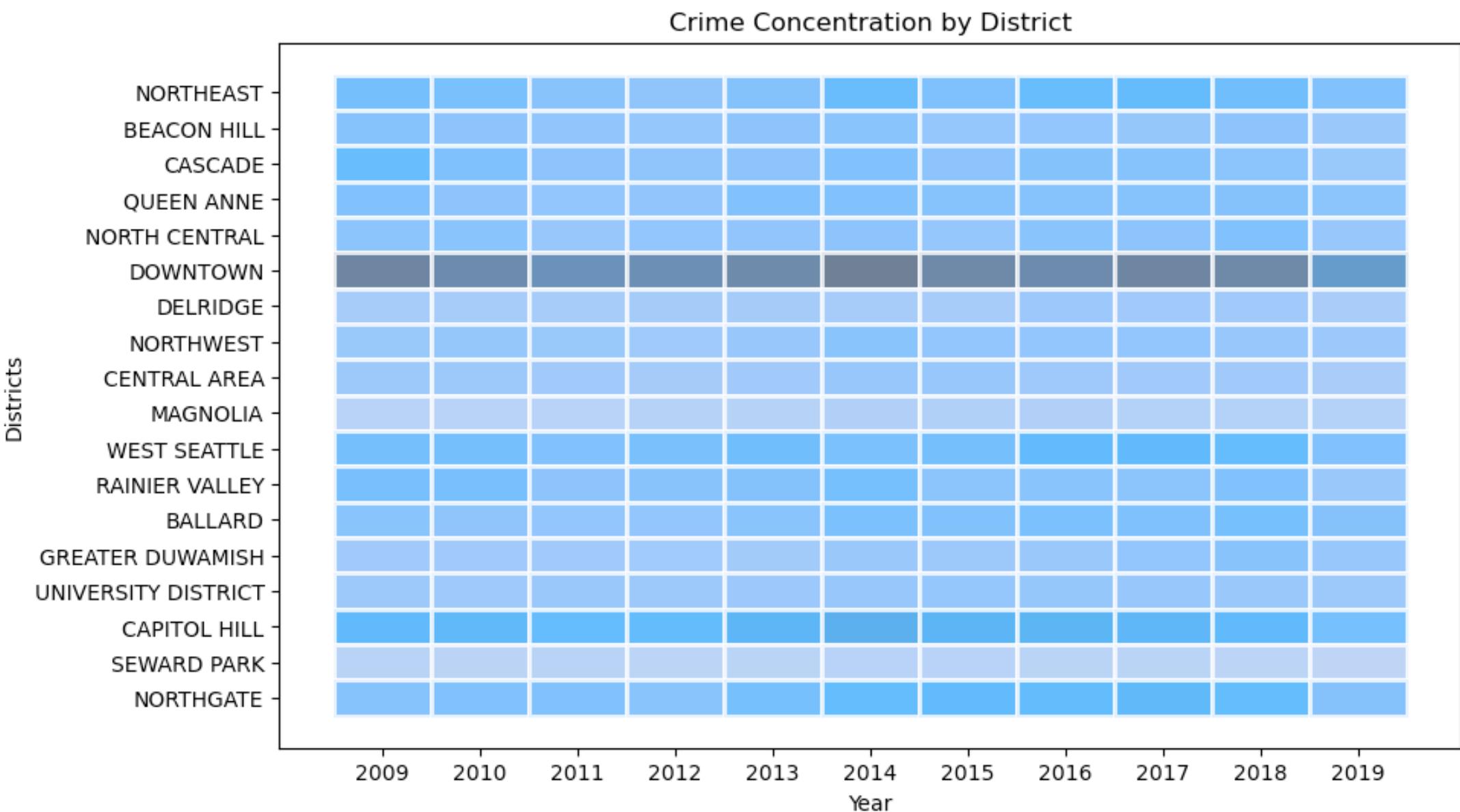


# EDA



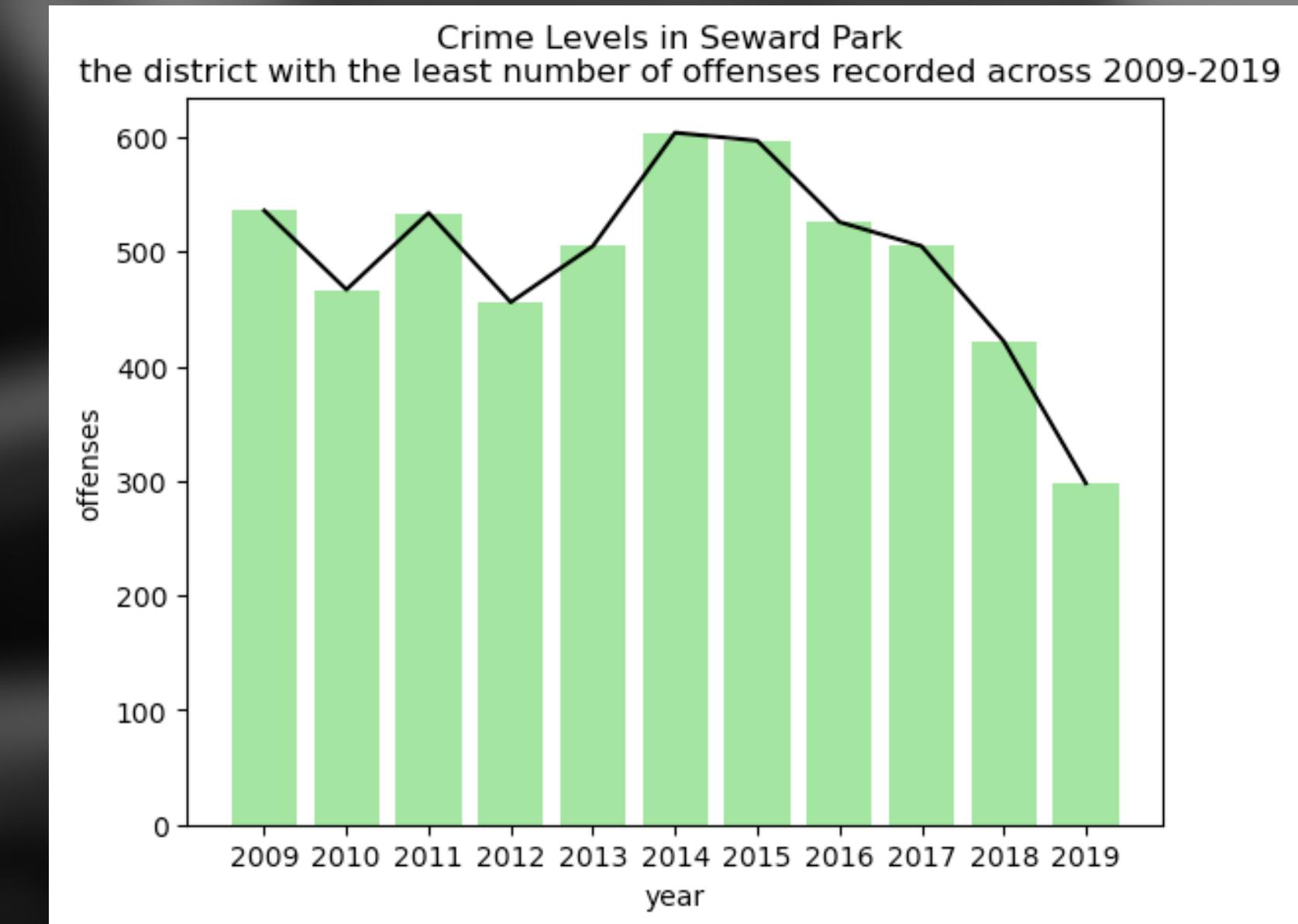
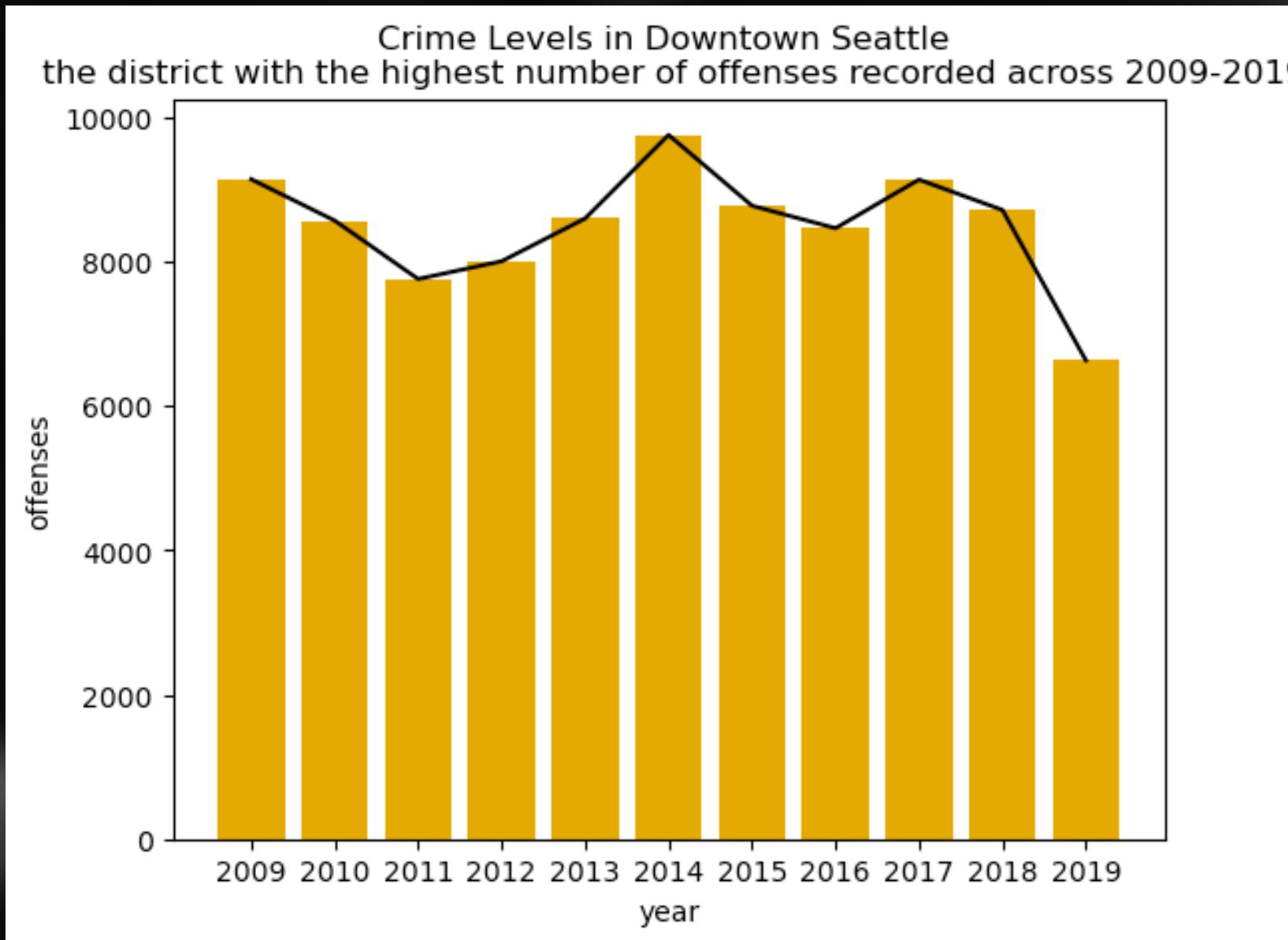
I created a violin plot to examine the general distribution of reported crimes in more detail. This visualization enabled me to observe that although there are numerous outliers for each year, they form a distinct subgroup within the dataset, generating a small but noticeable cluster of their own. Additionally, I noted that the median is not consistently positioned precisely in the center, and the overall shape of the violin plot demonstrates a similar trend. When the median leans towards the first quartile, it indicates a negative skew, suggesting a relative decrease in crime with a higher concentration of values smaller than the median.

# EDA



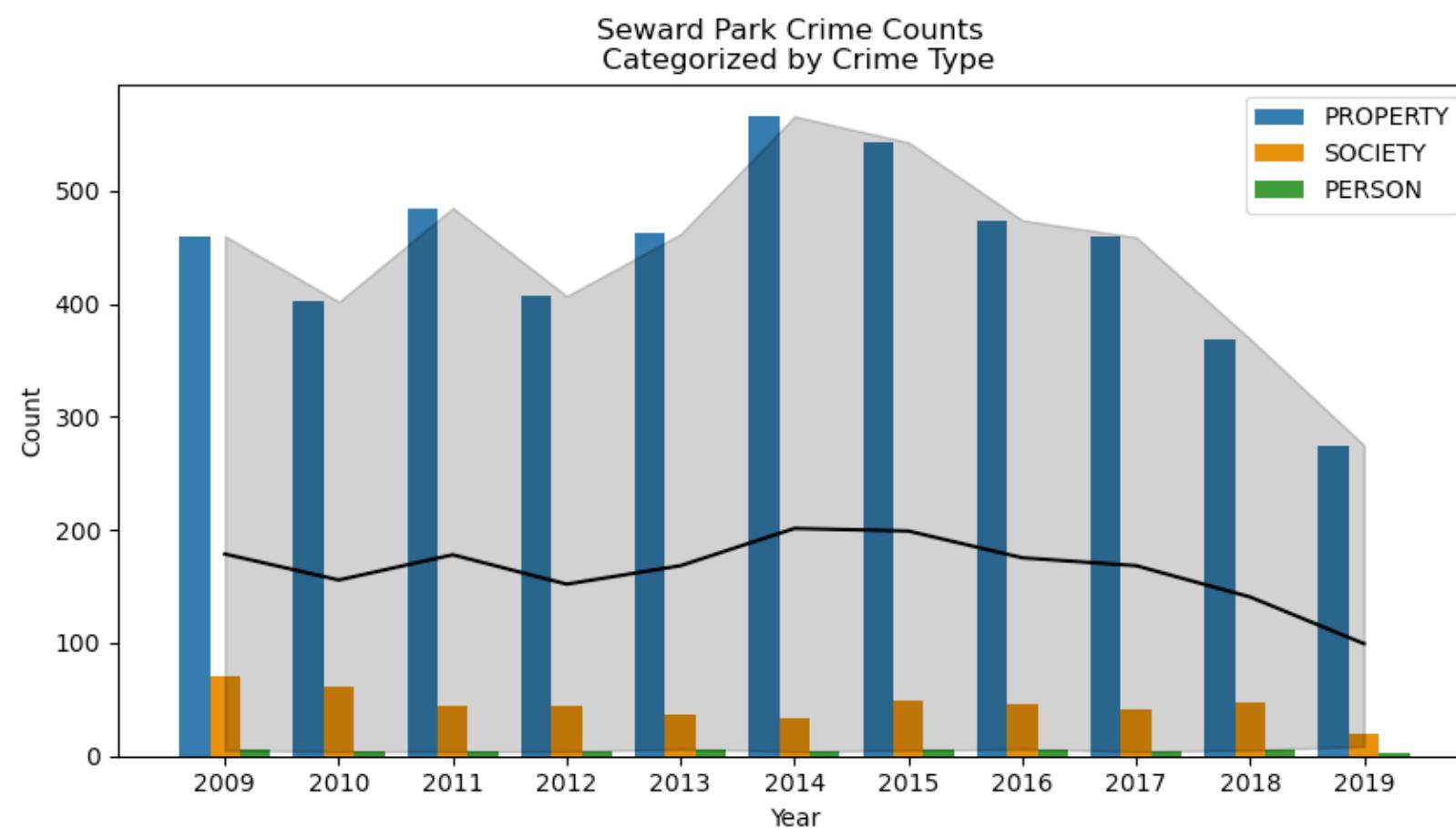
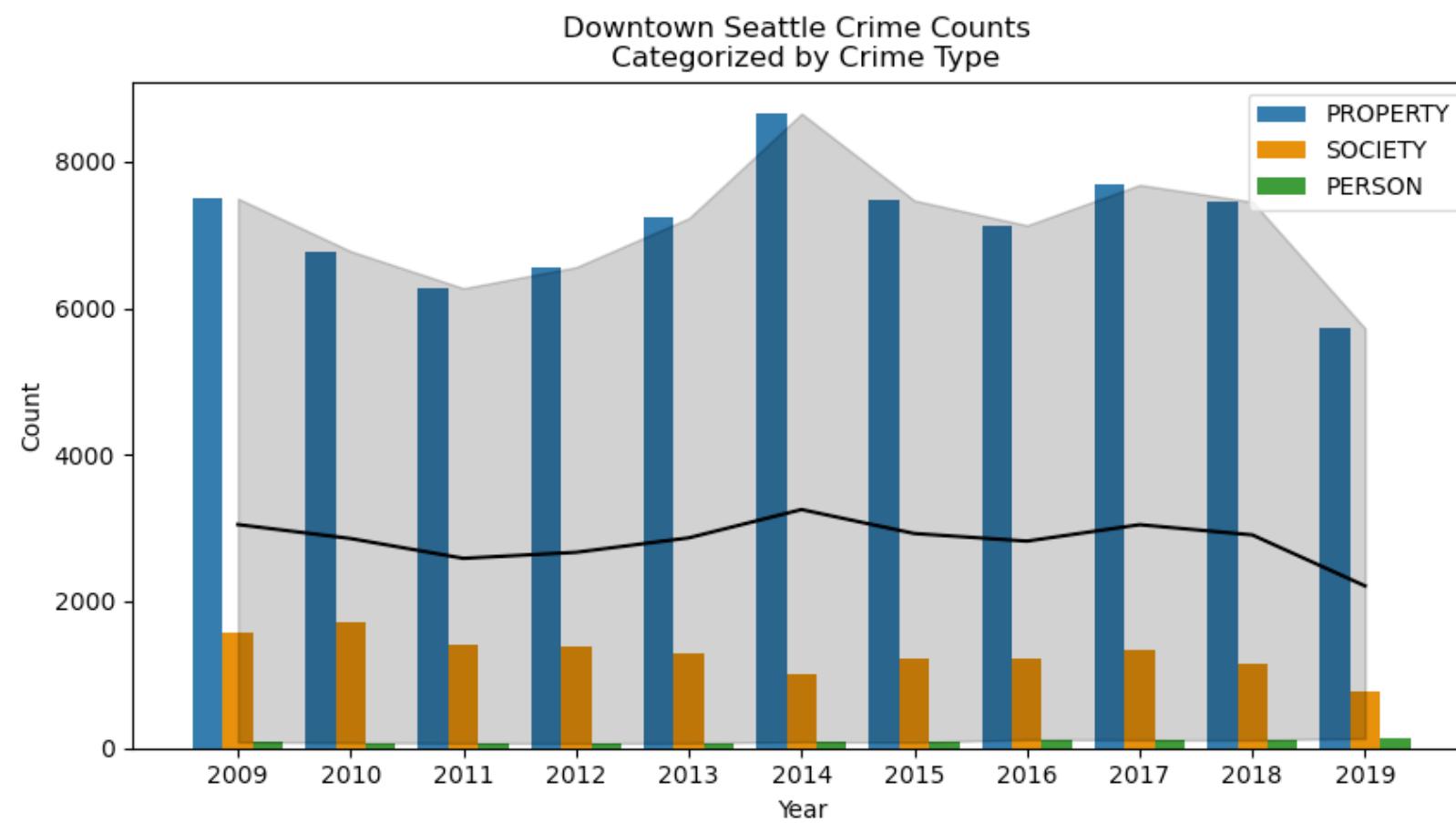
After identifying a subgroup in the data through the violin plot analysis, I proceeded to plot the distribution of these recorded offenses across the various districts to pinpoint the group responsible for the smaller cluster. Notably, one district appeared to have a significantly higher number of observations, which was expected. However, it was surprising to observe two districts exhibiting a consistent trend with the least number of observations.

# Analysis



Although the Downtown district saw a lot more crime than Seward Park, both followed the general crime trend line, seeing a drop from 2018 to 2019, and an increase between 2013 and 2015.

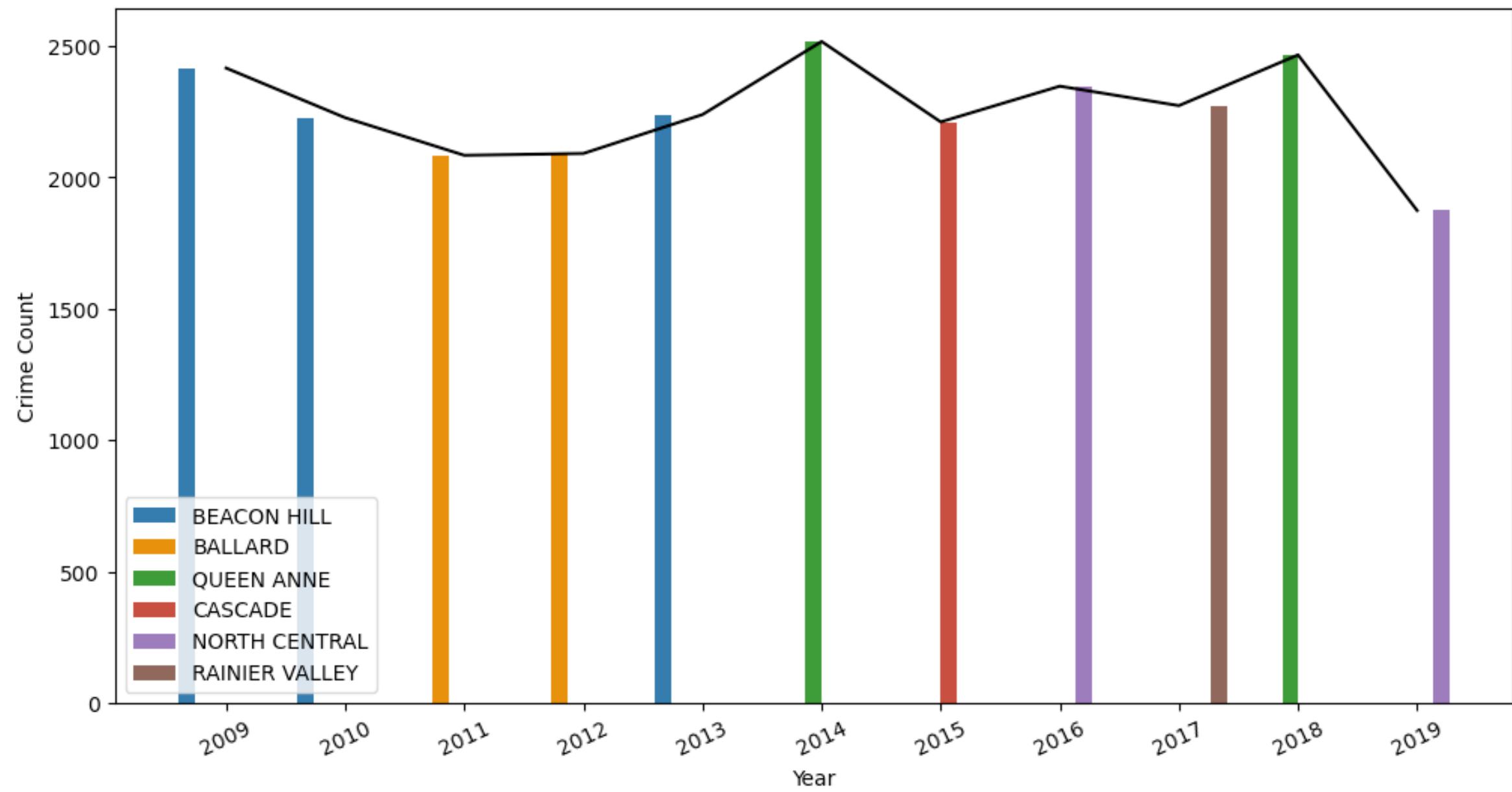
# Analysis



After a thorough examination of both districts, I found that they exhibit comparable distributions across property, person, and society crimes. The lines in both plots provide insight into the average distribution of all crime types and highlight how the property category predominantly influences the upward trend. The tops of the shaded areas mirror the previous lines observed in slide 8.

# Analysis

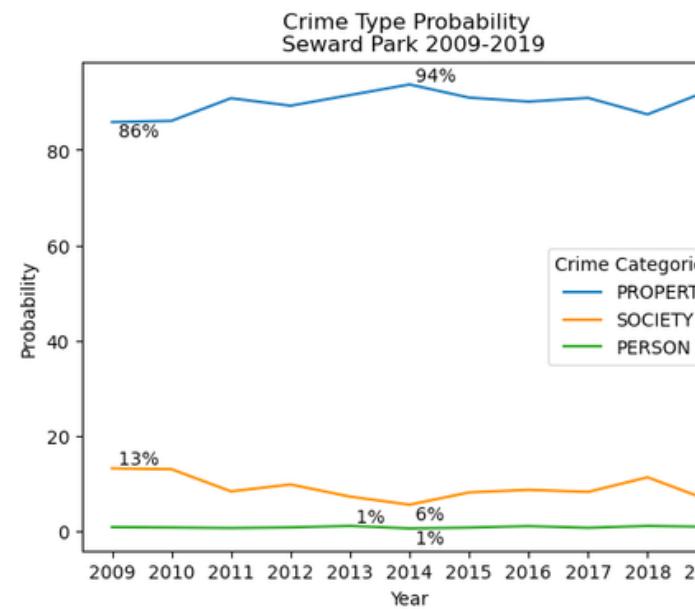
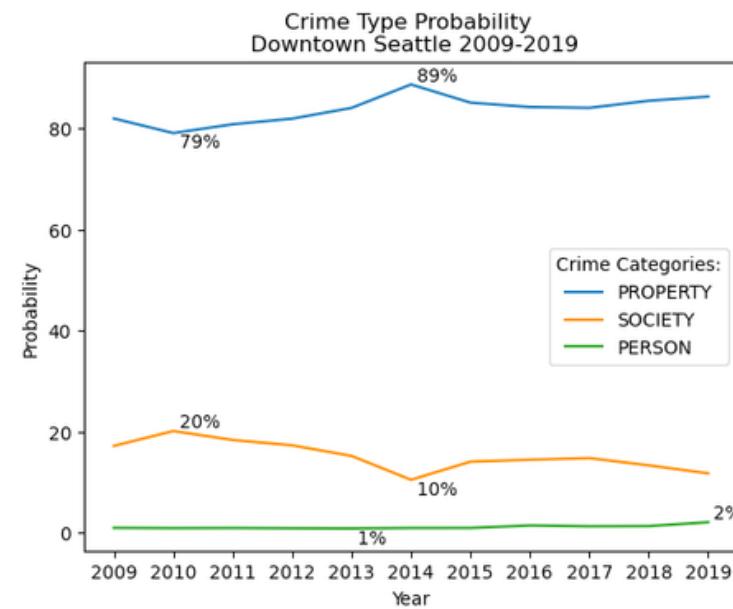
Mid-level Crime Across 6 Districts  
does the middle remain the same across 2009-2019?



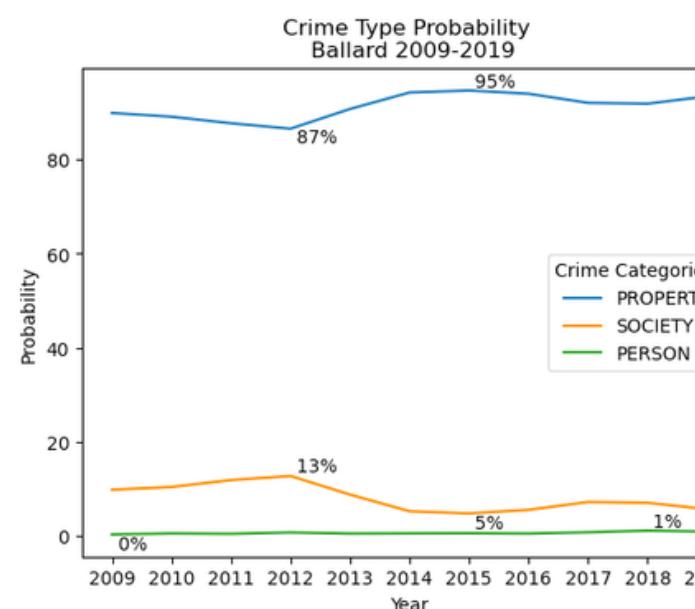
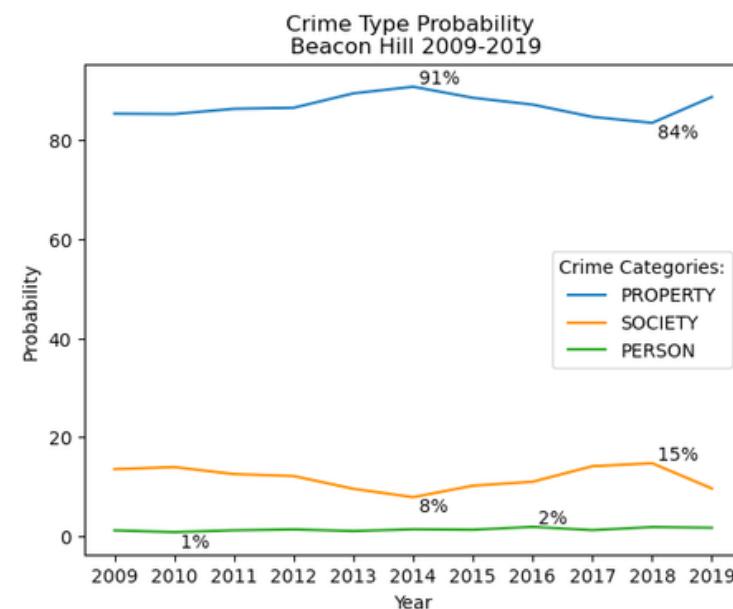
Another intriguing question that emerged during the exploratory data analysis phase of this project was whether a specific district consistently housed the median or 50th percentile of reported crimes, akin to Downtown's prominence in hosting the highest number of crime reports. It turned out that the mid-level crime districts varied, with six different ones identified over the decade of interest. Notably, only a couple of districts repeated twice in a row, namely Beacon Hill and Ballard.

# Analysis

Probability of a specific type of crime being reported across Downtown, Seward Park, Beacon Hill, and Ballard.



Although property crime remains the predominant type of crime in the city, the likelihood of encountering this type of crime in Downtown is marginally lower compared to Seward Park, Beacon Hill, or Ballard.



Another intriguing observation from these visualizations is the negative correlation between property and society crimes. As one decreases, the other tends to increase across these four districts.

# Interest

As a new resident in the Emerald City, I have heard plenty of stories about the perceived lack of safety in the metro area. After moving here in December of 2022, I found news stories often employed hyperbole in service of viewership, detracting from the underlying root causes yielding to feelings of insecurity.

While I wanted to conducted a more recent analysis on crime to account for and assess the pandemic's impact during and after lockdown, I encountered a data provenance issue. Although the data was purported to be updated daily, I quickly learned SPD (Seattle Police Department) revamped their record system in 2020, which was supported by the date of the last entry on this dataset (Feb. 2020).

Therefore, I redirected my focus to the decade preceding COVID to gain a comprehensive understanding of the metro crime trends and lay the groundwork for a future project.

# Next steps

This project has effectively debunked the misconception of Seattle being an exceptionally dangerous city. However, I'm keen to delve deeper into how different demographics and backgrounds influence individuals' personal encounters with crime. Just as not all crimes are alike, considering demographics can offer further insights into residents' perspectives and experiences within the city.

Furthermore, with a decade's worth of data at my disposal, I intend to perform logistic regression to forecast how crime in Seattle might have unfolded in 2020 had the pandemic not occurred. Despite the Seattle Police Department (SPD) making data from 2020 onwards less accessible, I aim to compare their dashboard visualizations with my predictions to gauge the pandemic's impact.





**THANK  
YOU**

