

# Police Data Challenge

## The Bayes

Dave Ho — Santosh Tatipamula — Alexander Vu — Luke Liu — Tyler Shockley

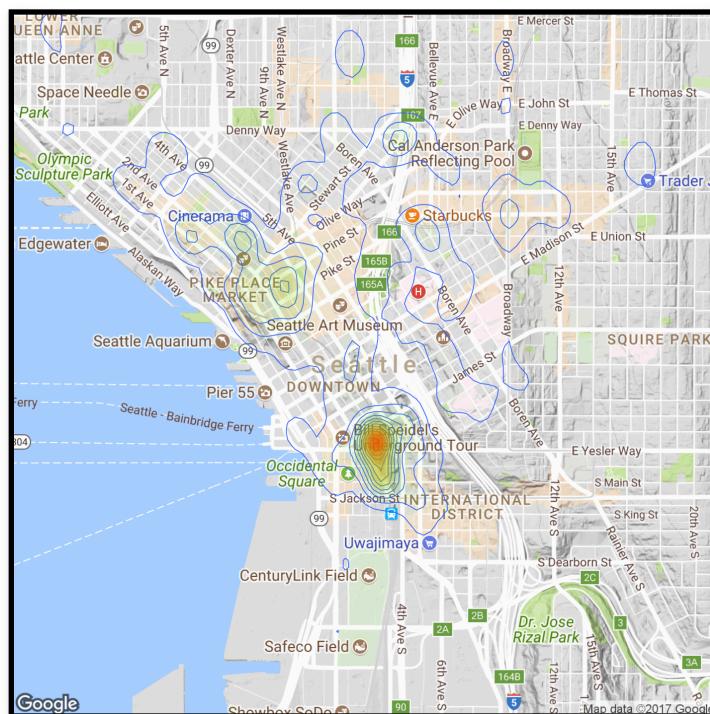


# Introduction

As technology continues to advance, the analysis of data is slowly becoming an essential part of operations across all industries. Discovering valuable patterns to help allocate future resources and improve working efficiency is vital in the public sector, even more specifically in law enforcement. The data generated by law enforcement in their daily operations can be analyzed with statistical tools to provide a myriad of benefits, including a safer environment for law enforcement, improvement of neighborhood safety, and promotion of community outreach, to name a few. Our goals during this challenge were as follows: analyze the existing data from the Seattle Police Department and develop recommendations to improve public safety.

## Brainstorming

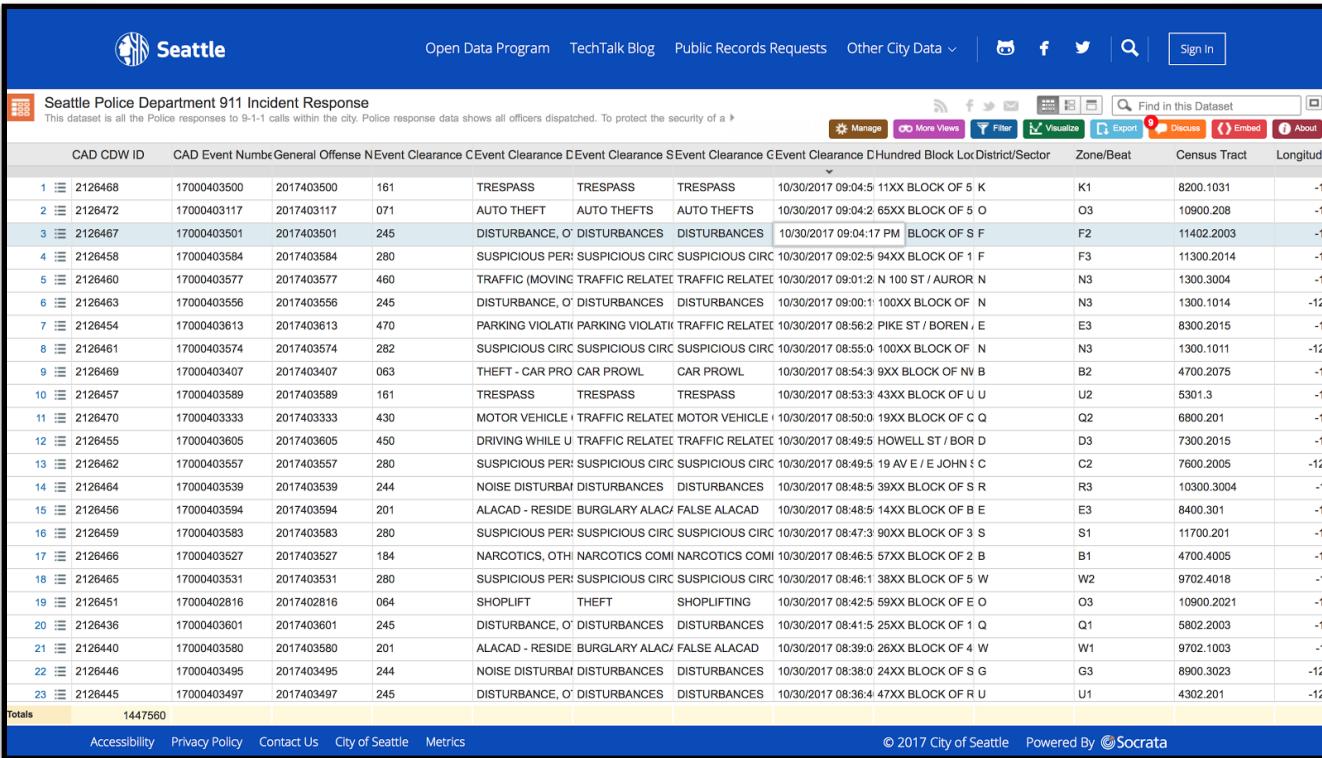
Initially, our group looked at fire data. Our thinking was by observing the density of fires within certain areas of Seattle we can find areas that needed more fire prevention and firefighting infrastructure. Utilizing a combination of Google Maps API and R Software we were able to generate a heat map of fires across Seattle.



After analyzing the graph of fires in the City of Seattle, we discovered that there was a much larger concentration of fires near the downtown area than anywhere else. More specifically near the north end of Chinatown-International District. However helpful the results towards the Fire Department, our group wanted to apply our analysis more to the City of Seattle's Policing Force, as this is the "Police Data Challenge".

## Collecting and Cleaning Data

We briefly glanced at the data recorded on the Seattle Gov website. After peering into several datasets, we examined the Police Incident Report table set and decided to utilize it for our analysis.

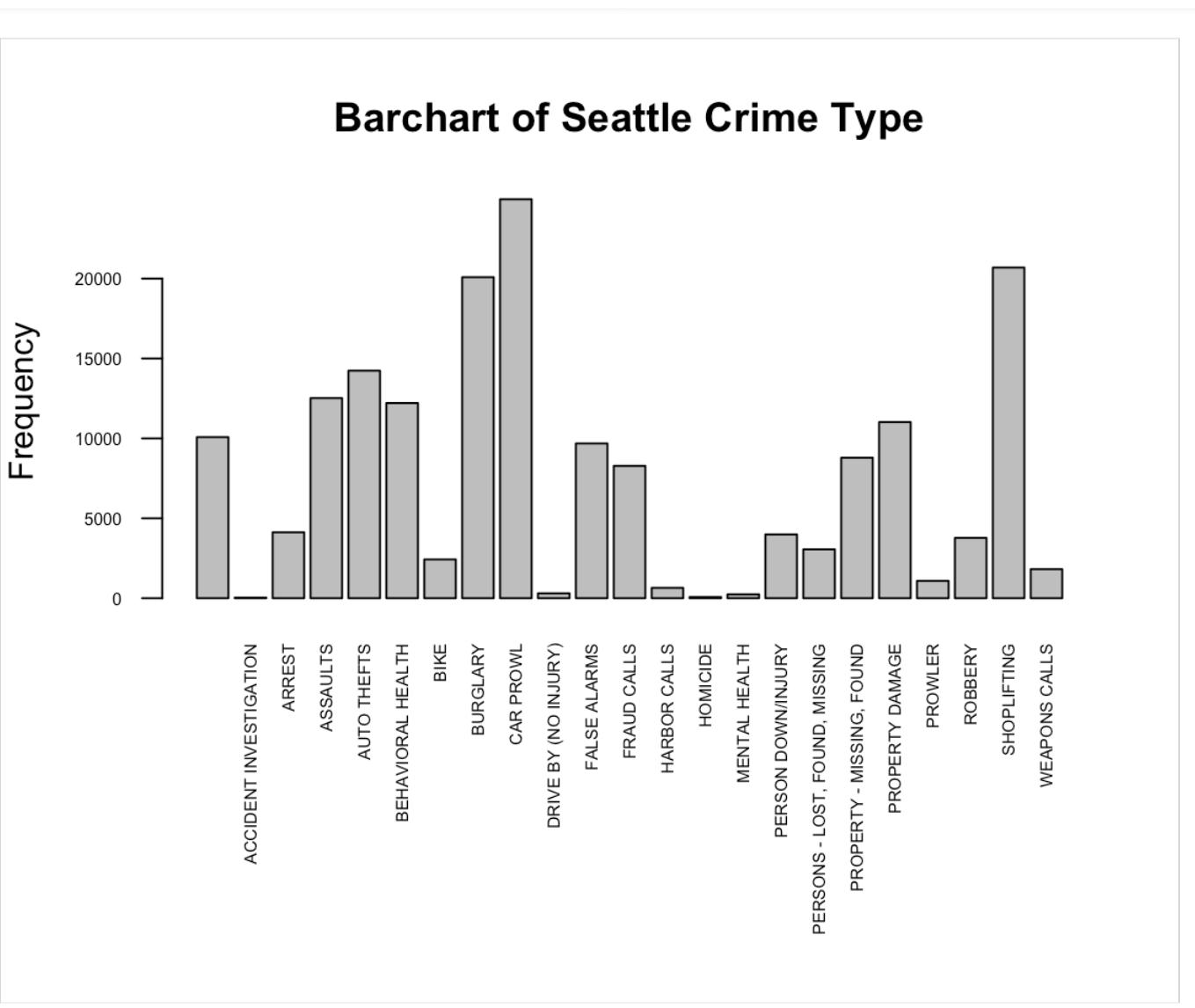


The screenshot shows a table titled "Seattle Police Department 911 Incident Response". The table has 17 columns: CAD CDW ID, CAD Event Numbr, General Offense NEvent Clearance CEvent Clearance DEvent Clearance SEvent Clearance GEvent Clearance CHundred Block Loc, District/Sector, Zone/Beat, Census Tract, and Longitude. The table contains approximately 1.4 million rows of data. The header row includes column names and descriptions. The footer row shows "Totals 1447560". At the bottom, there are links for Accessibility, Privacy Policy, Contact Us, City of Seattle, and Metrics. The page is powered by Socrata.

CAD CDW ID	CAD Event Numbr	General Offense NEvent Clearance CEvent Clearance DEvent Clearance SEvent Clearance GEvent Clearance CHundred Block Loc	District/Sector	Zone/Beat	Census Tract	Longitude
1	2126468	17000403500	2017403500	161	TRESPASS TRESPASS TRESPASS	10/30/2017 09:04:5 11XX BLOCK OF 5 K
2	2126472	17000403117	2017403117	071	AUTO THEFT AUTO THEFTS AUTO THEFTS	10/30/2017 09:04:2 65XX BLOCK OF 5 O
3	2126467	17000403501	2017403501	245	DISTURBANCE, O' DISTURBANCES DISTURBANCES	10/30/2017 09:04:17 PM BLOCK OF S F
4	2126458	17000403584	2017403584	280	SUSPICIOUS PER: SUSPICIOUS CIRC SUSPICIOUS CIRC SUSPICIOUS CIRC	10/30/2017 09:02:5 94XX BLOCK OF 1 F
5	2126460	17000403577	2017403577	460	TRAFFIC (MOVING) TRAFFIC RELATED TRAFFIC RELATED	10/30/2017 09:01:2 N 100 ST / AUROR N
6	2126463	17000403556	2017403556	245	DISTURBANCE, O' DISTURBANCES DISTURBANCES	10/30/2017 09:00:1 100XX BLOCK OF N
7	2126454	17000403613	2017403613	470	PARKING VIOLATI PARKING VIOLATI TRAFFIC RELATED	10/30/2017 08:56:2 PIKE ST / BOREN , E
8	2126461	17000403574	2017403574	282	SUSPICIOUS CIRC SUSPICIOUS CIRC SUSPICIOUS CIRC SUSPICIOUS CIRC	10/30/2017 08:55:0 100XX BLOCK OF N
9	2126469	17000403407	2017403407	063	THEFT - CAR PRO CAR PROWL CAR PROWL	10/30/2017 08:54:3 9XX BLOCK OF NW B
10	2126457	17000403589	2017403589	161	TRESPASS TRESPASS TRESPASS	10/30/2017 08:53:3 43XX BLOCK OF U U
11	2126470	17000403333	2017403333	430	MOTOR VEHICLE I TRAFFIC RELATED MOTOR VEHICLE I	10/30/2017 08:50:0 19XX BLOCK OF C Q
12	2126455	17000403605	2017403605	450	DRIVING WHILE U TRAFFIC RELATED TRAFFIC RELATED	10/30/2017 08:49:5 HOWELL ST / BOR D
13	2126462	17000403557	2017403557	280	SUSPICIOUS PER: SUSPICIOUS CIRC SUSPICIOUS CIRC SUSPICIOUS CIRC SUSPICIOUS CIRC	10/30/2017 08:49:5 19 AV E / JOHN S C
14	2126464	17000403539	2017403539	244	NOISE DISTURBA DISTURBANCES DISTURBANCES	10/30/2017 08:48:5 39XX BLOCK OF S R
15	2126456	17000403594	2017403594	201	ALACAD - RESIDE BURGLARY ALAC/ FALSE ALACAD	10/30/2017 08:48:5 14XX BLOCK OF B E
16	2126459	17000403583	2017403583	280	SUSPICIOUS PER: SUSPICIOUS CIRC SUSPICIOUS CIRC SUSPICIOUS CIRC SUSPICIOUS CIRC	10/30/2017 08:47:3 90XX BLOCK OF 3 S
17	2126466	17000403527	2017403527	184	NARCOTICS, OTH/ NARCOTICS COMI NARCOTICS COMI	10/30/2017 08:46:5 57XX BLOCK OF 2 B
18	2126465	17000403531	2017403531	280	SUSPICIOUS PER: SUSPICIOUS CIRC SUSPICIOUS CIRC SUSPICIOUS CIRC SUSPICIOUS CIRC	10/30/2017 08:46:1 38XX BLOCK OF 5 W
19	2126451	17000402816	2017402816	064	SHOPLIFT THEFT SHOPLIFTING	10/30/2017 08:42:5 59XX BLOCK OF E O
20	2126436	17000403601	2017403601	245	DISTURBANCE, O' DISTURBANCES DISTURBANCES	10/30/2017 08:41:5 25XX BLOCK OF 1 Q
21	2126440	17000403580	2017403580	201	ALACAD - RESIDE BURGLARY ALAC/ FALSE ALACAD	10/30/2017 08:39:0 26XX BLOCK OF 4 W
22	2126446	17000403495	2017403495	244	NOISE DISTURBA DISTURBANCES DISTURBANCES	10/30/2017 08:38:0 24XX BLOCK OF S G
23	2126445	17000403497	2017403497	245	DISTURBANCE, O' DISTURBANCES DISTURBANCES	10/30/2017 08:36:4 47XX BLOCK OF R U

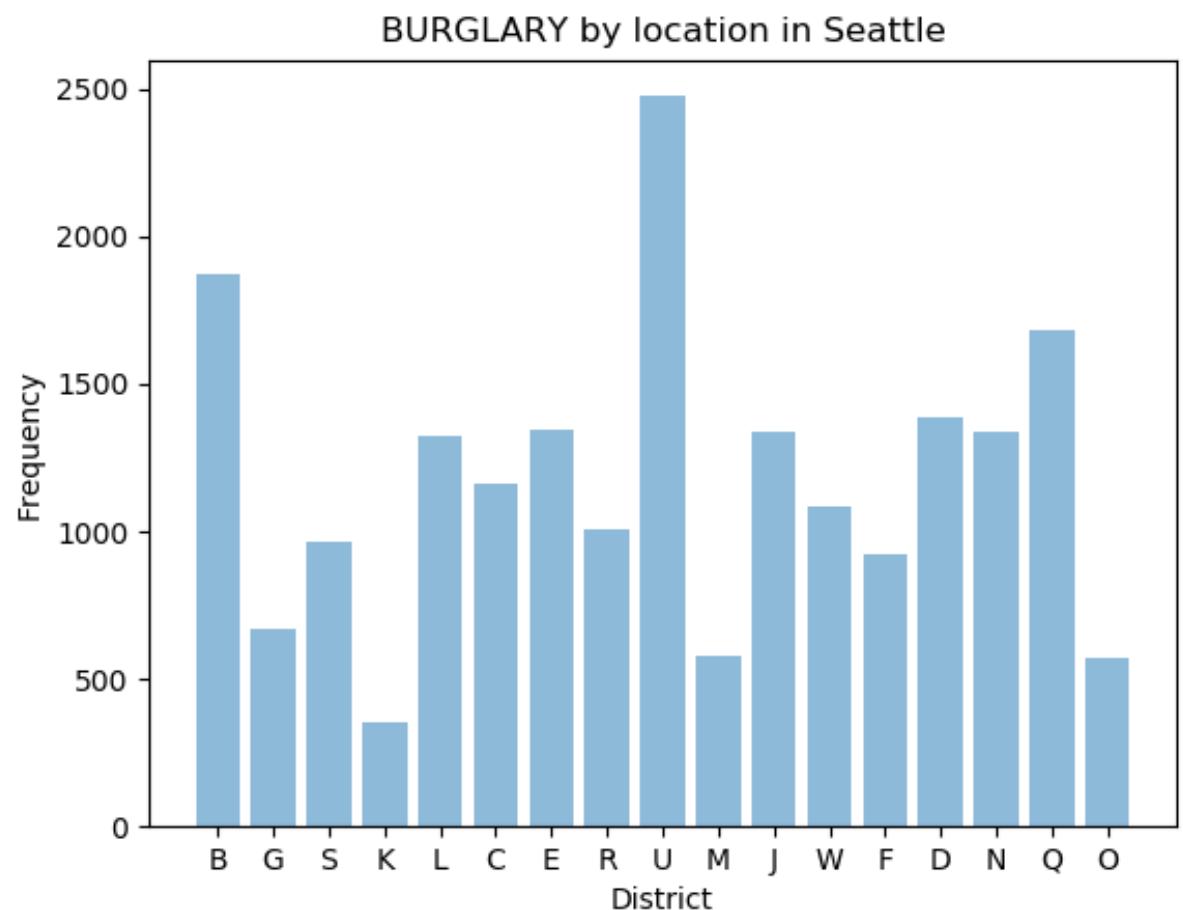
With approximately 1.5 million rows of data spanning from 2009 to 2017, we were both astonished by the sheer detail of the entries, as well as frustrated over the amount of "dirty data" mixed in with golden shards of well categorized rows. To solve the problem of dealing with large datasets as well as cleaning our data, we referred to the open source statistical software, R. Cutting down the datasets by removing the rows with missing data,

deleting 10 columns that were irrelevant to our analysis, and narrowing the target time period to 2009 to 2014, we were finally able to dive in the core part of our project. After much debate and difficulty, our group concluded that a total summary of all crimes was needed to see what aspect to focus on. The three most common crimes in Seattle were as follows, in order from most to least common: Car Prowling, Shoplifting, and Burglary. By splitting these crimes we could better visualize the criminal world of Seattle.

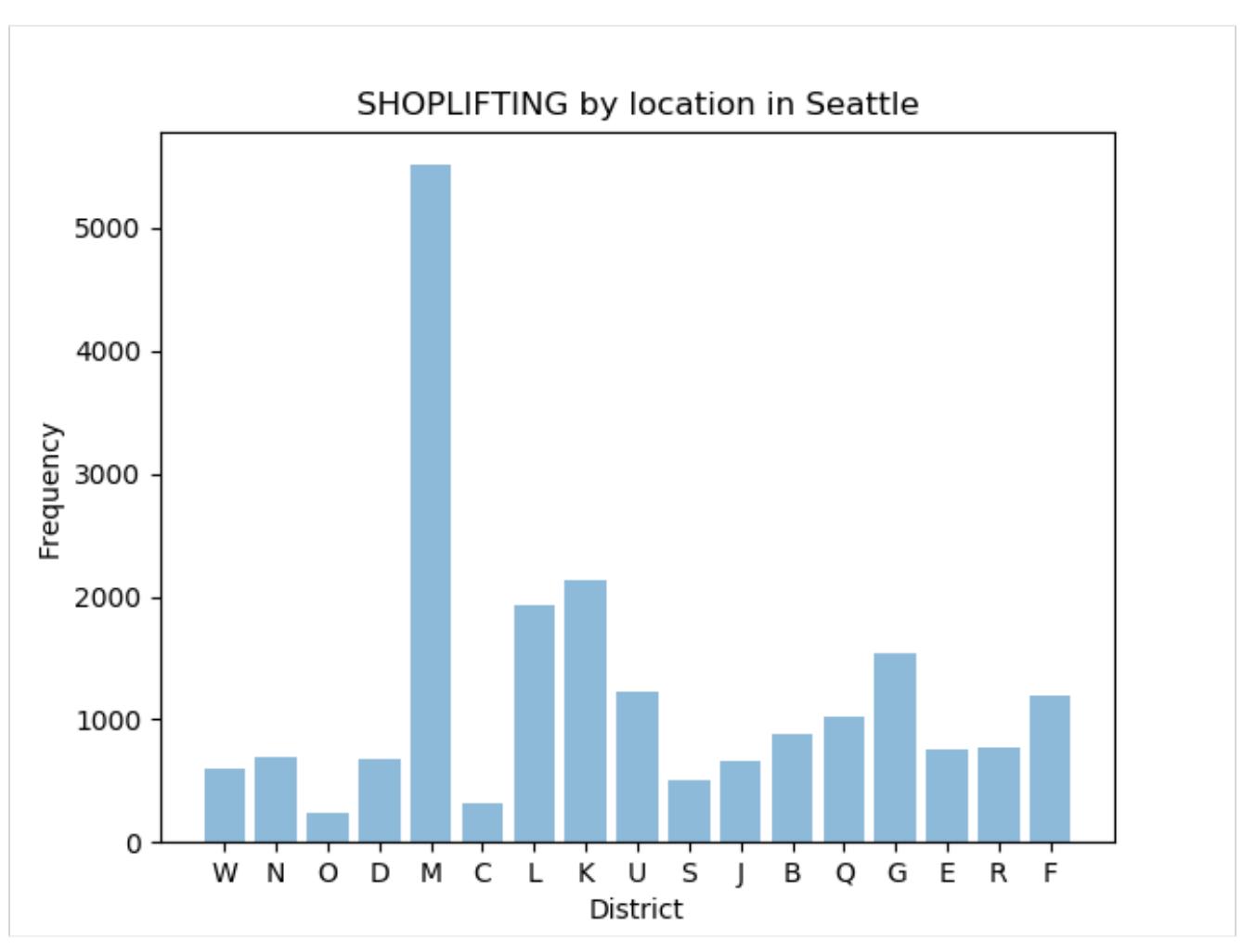


## Data Report

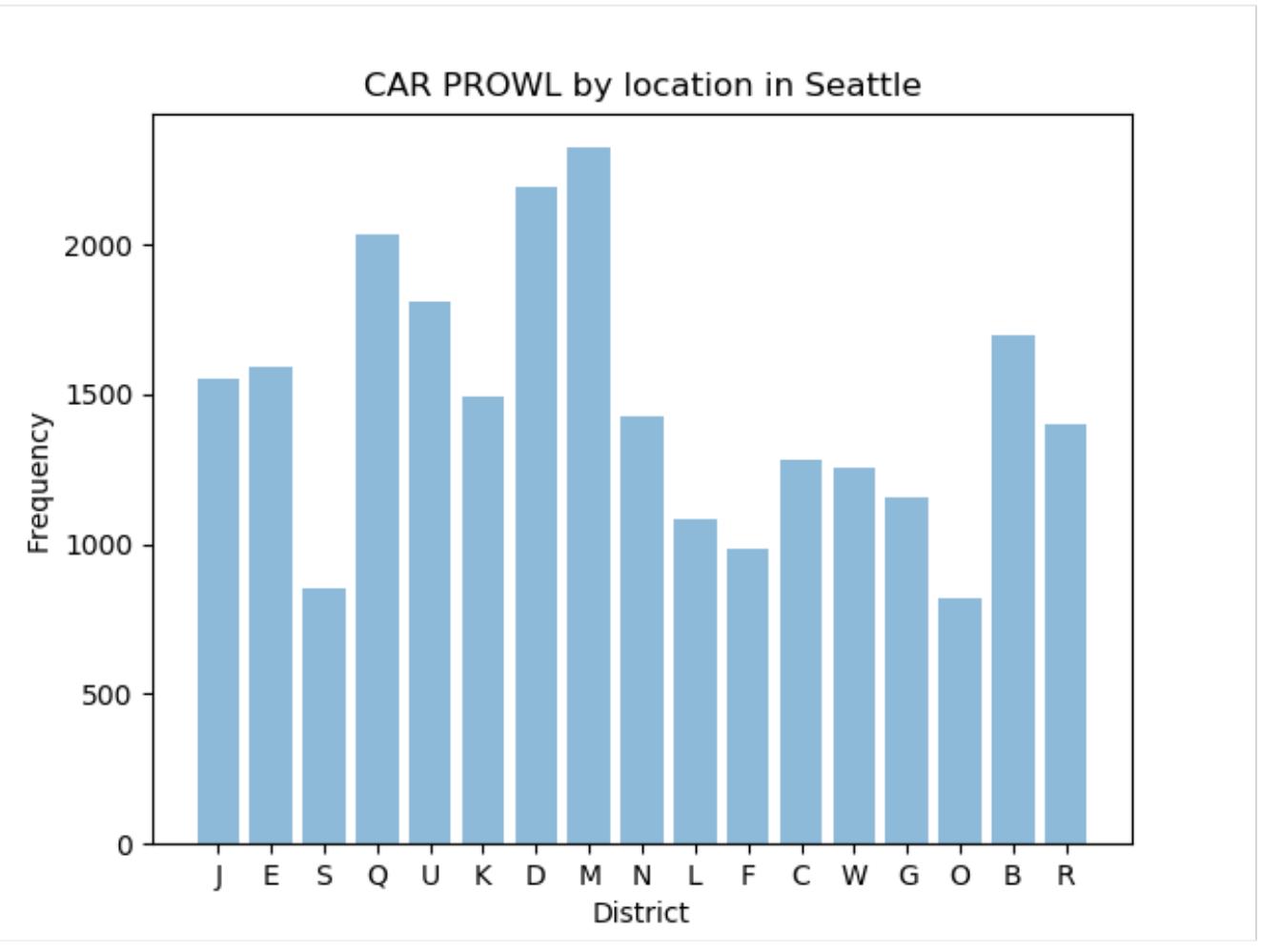
Seattle is mapped into 17 different policing districts, each represented by a letter. Via analyzing the frequency of a certain crime in each district, we can find concentrated points of specific criminal activity. Burglary, the criminal offense of breaking and entering a building illegally for the purpose of committing a crime, is most commonly reported in district U and B, both of which are residential areas.



Shoplifting, the willful theft of merchandise from a retail establishment without the knowledge or consent of the seller, is concentrated near Downtown Seattle, commercial M district.

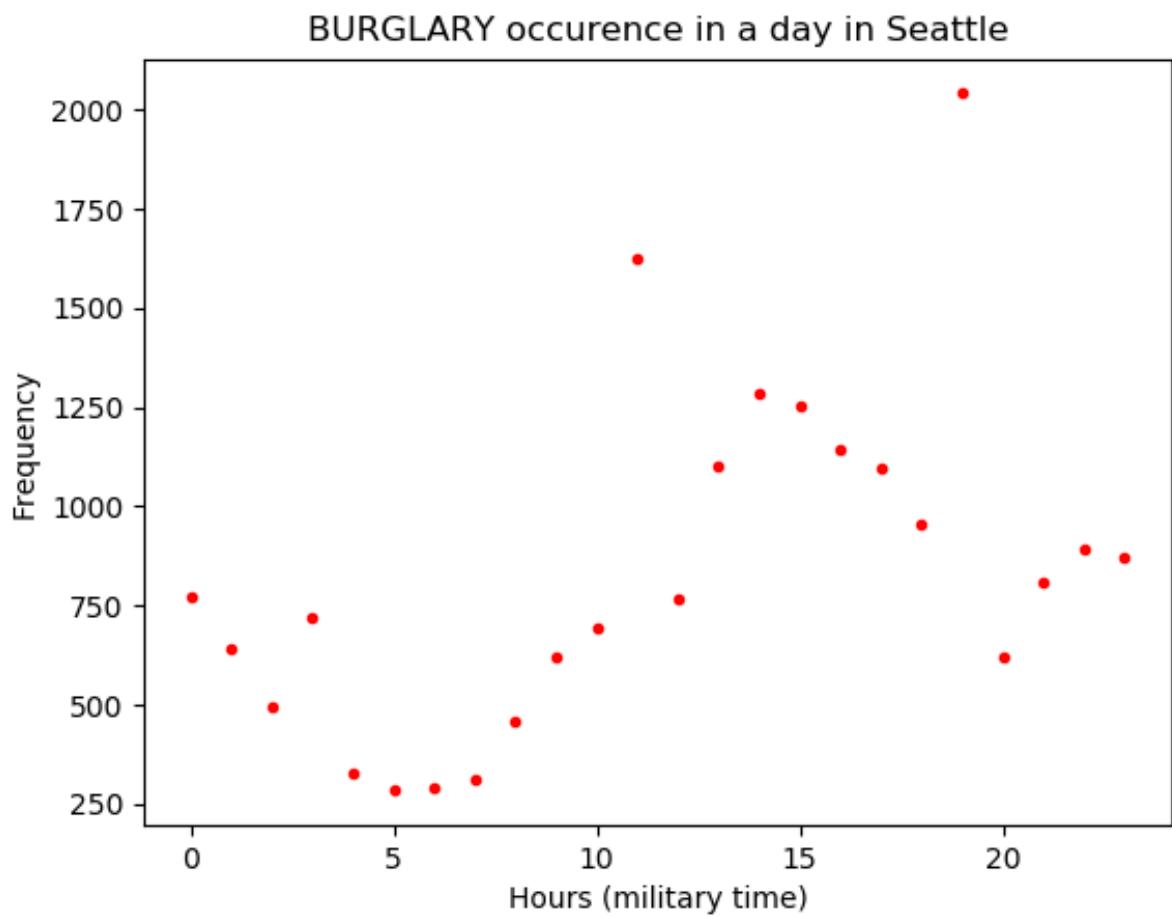


However, we found an interesting anomaly in car prowling, as the results show that car prowling is extremely common in almost seven districts. But, given the nature of stealing car parts from parked cars, it isn't surprising.

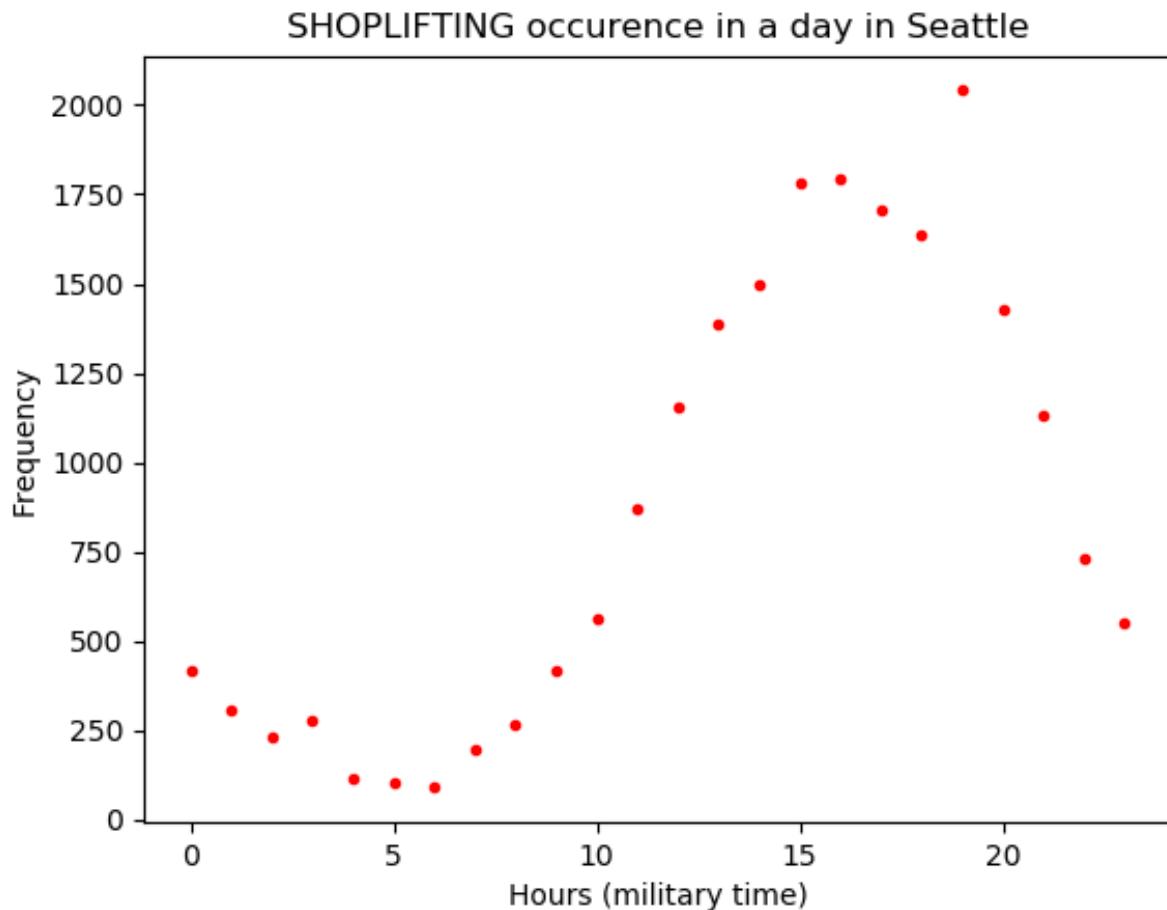


From 2015 to 2017, criminal activity spikes every few months. In fact, we noticed that each crime shows a jump every time July rolls around, with the exception of homicides. Without concluding anything, our group theorized that with summer vacation, troubled and bored youth attempt petty crimes. However, the data does not contain a suspect's profile, so further analysis with more information is required.

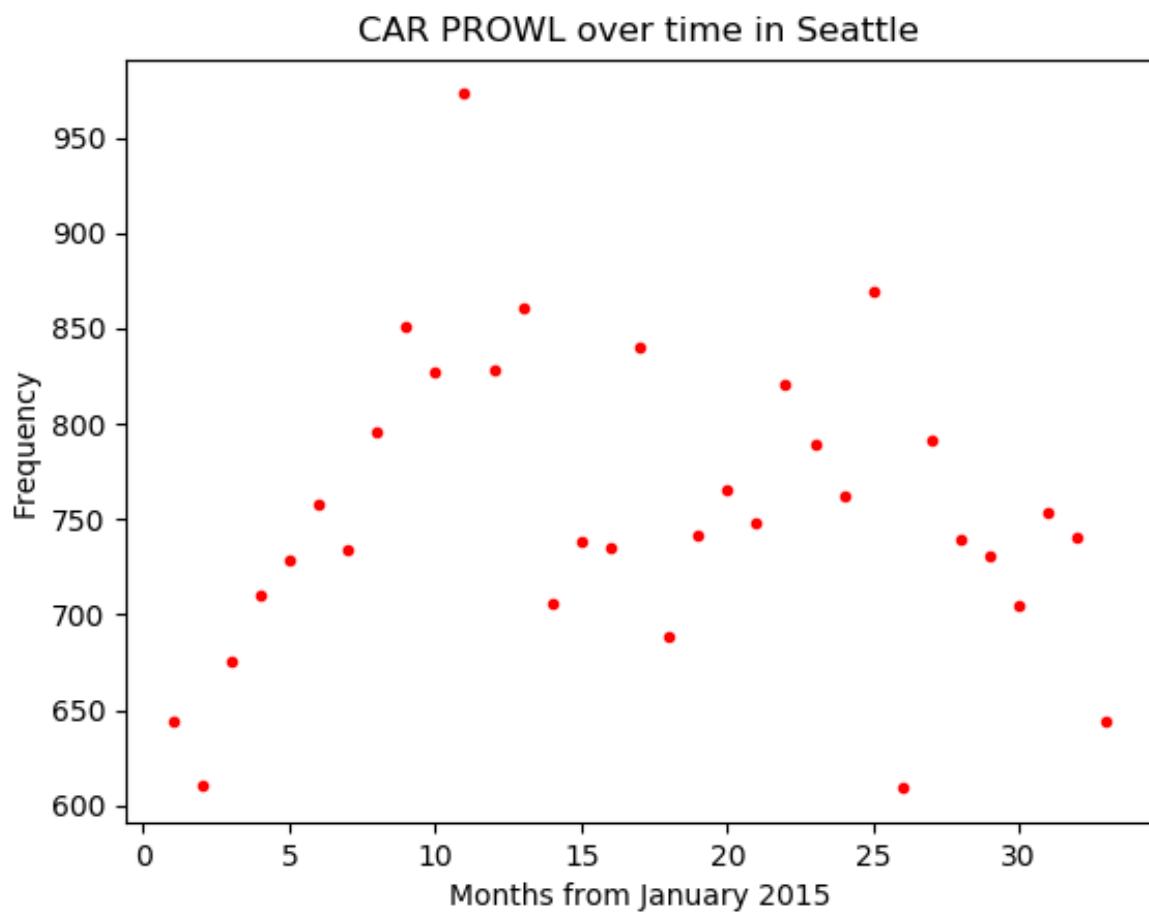
By breaking down specific times of day, our group found we could better visualize the distribution of crimes by the time they take place. Burglary, as expected, mostly occurs at about mid afternoon and in late evening to early morning, times which either a homeowner is away or asleep, respectively.



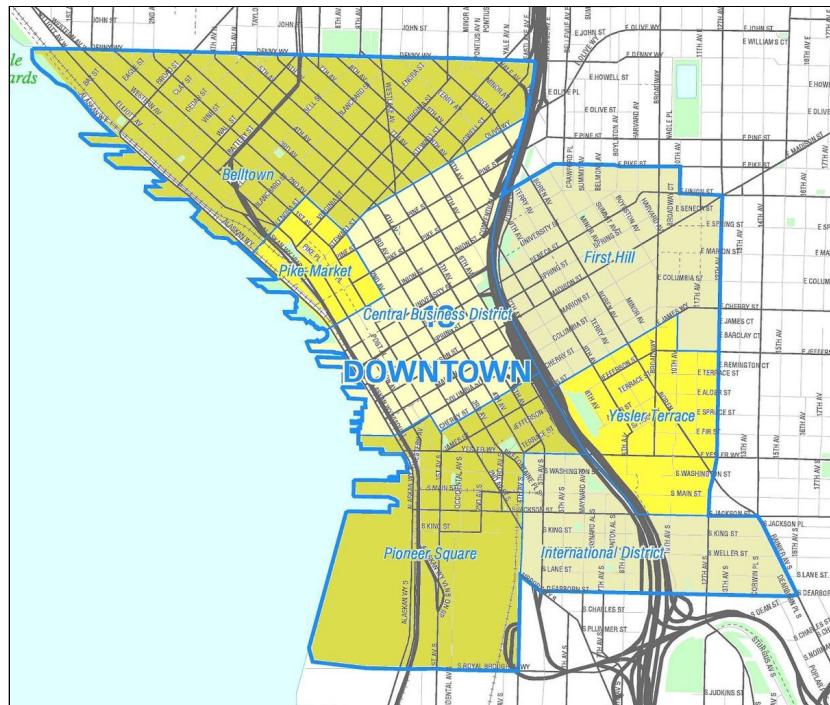
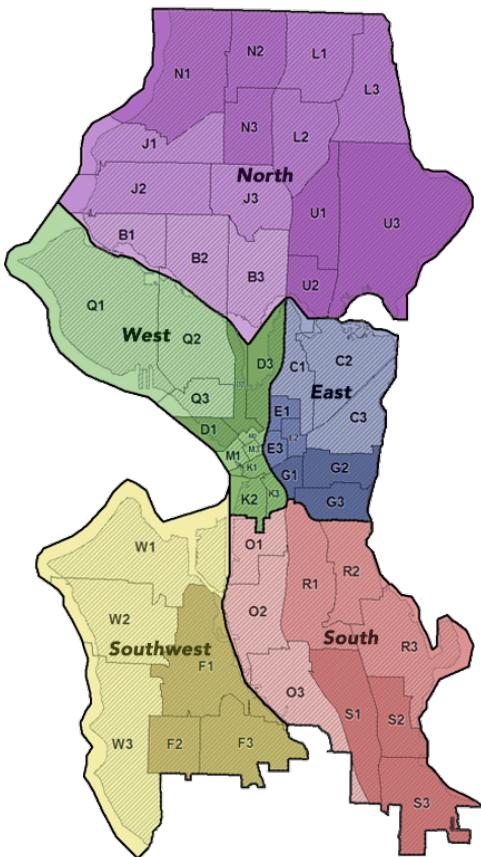
Shoplifting sees a sharp increase as late morning arrives, peaking at about three o'clock in the afternoon. This is about the time where many stores are most active in business and vulnerable to slight handed individuals.



Car prowling is the most interesting of the three crimes analyzed, as its most popular times are time periods which many would expect to have the largest amount of witnesses to any illicit endeavors. From seven in the morning to late evening is not really the perfect time for such a crime, although, the action of car prowling hinges upon the quick disassembly of a parked car's parts, and as such, should be done in daylight to reduce time spent in action (taken from experience in vehicle repair).



## Proposals



Based on the analysis of the data provided, our group has come up with several recommendations. The M District is one of the most popular locations for shoplifting, and as a result, should be better patrolled by officers. Specifically the area with heavy occurrences of shoplifting is the infamous Pike Market, a public market overlooking the Elliott Bay waterfront in Seattle. Public Markets, like these, could be extremely crowded, especially during busy hours, providing very limited room for cars. Thus, our group proposes police to increase usage of bikes or motorcycles as a means of patrolling that area.



---

Furthermore, although shoplifting isn't a crime that can be easily caught, our group recommends the Seattle Police Department reach out to business owners on how to better prevent shoplifting through monitoring and surveillance. Burglary, as shown in previous data, is very popular in the U District. After conducting some research, we found out that this was where the University of Washington was located. We propose, campus police to enforce stricter rules and consequences to deter such crimes. In conclusion, the suggestion our group proposes is to build more community outreach programs for troubled suspects such as educational and employable skill training, something that we believe will better encourage these citizens to become more productive members of Seattle's thriving society.

---

## Sources

1. <https://www.r-project.org/>
2. [https://www.jmp.com/en\\_us/home.html](https://www.jmp.com/en_us/home.html)
3. <http://ggplot2.org/>
4. <https://matplotlib.org/>
5. <https://data.seattle.gov/>