Data Science Seminar - MSAI 339

Checkpoint 2

October 21st, 2021

Professor
Jennie Rogers

Students
Aleksandr Simonyan
Dimitrios Mavrofridis
Donald Baracskay

Introduction

For Checkpoint 2 of our final project, we are aspiring to answer our initial visualization questions by integrating the CPDP data and performing the appropriate tasks in order to generate quality data visualizations using Tableau. The main theme of our project revolves around the Tactical Response Reports. We intend to examine the areas and officers involved in violent incidents and whether officers listed in TRRs receive more awards than those who deal with less violent incidents or vice versa. Finally, we are going to determine the percentage of officers that are using excessive force when dealing with tactical responses. All these are represented by visuals and figures generated with Tableau based on the following chart types:

1) Packed Bubble Charts
2) Box & Whiskers Charts
3) Stacked Vertical Bar Graphs.

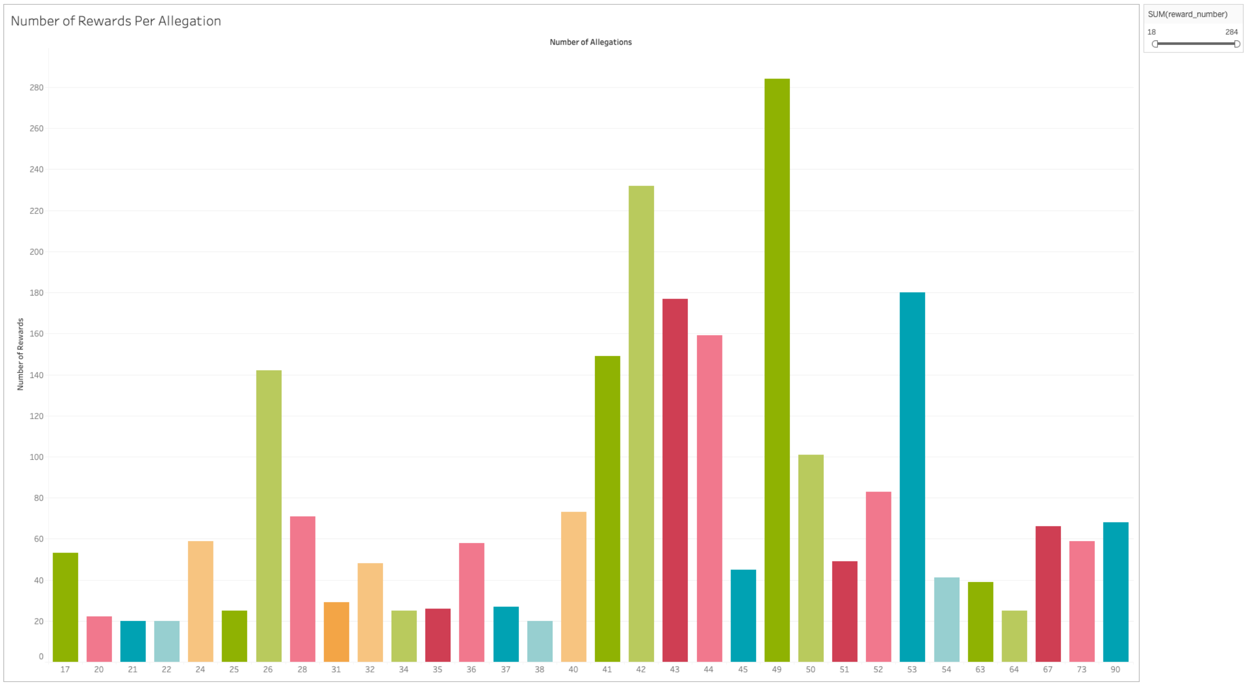
In the next section we are going to be discussing the questions in depth as well as analyzing the visual representations of our data.

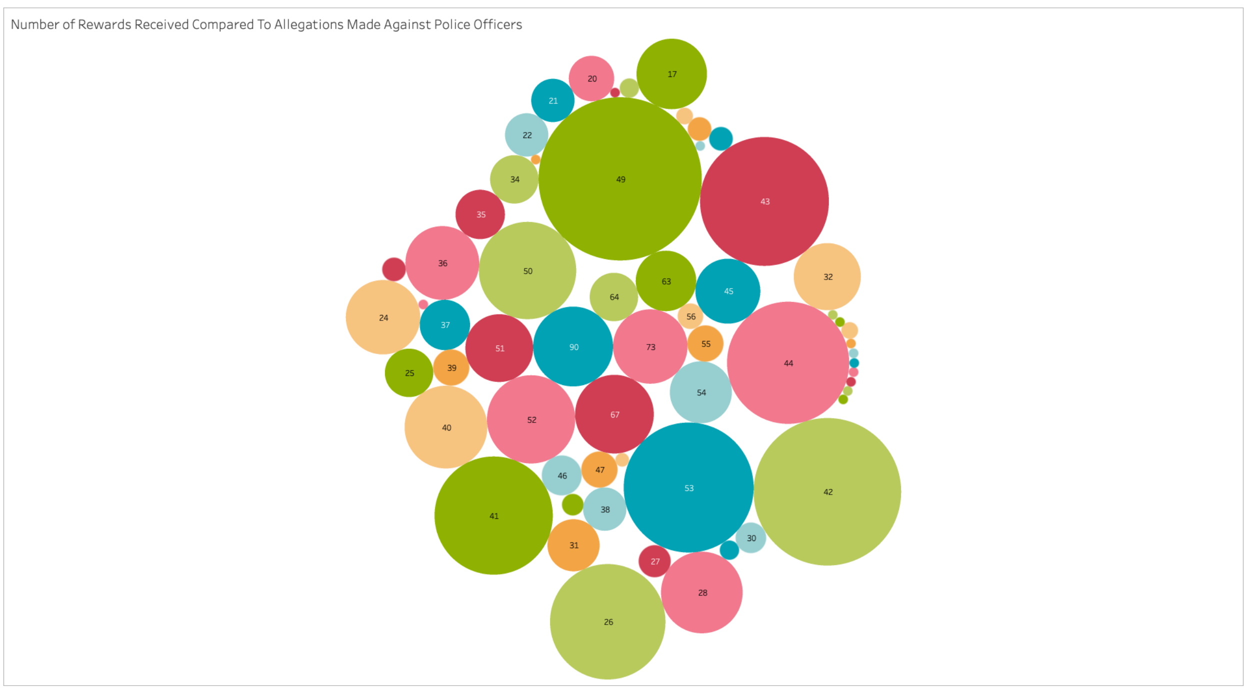
Findings & Results:

Question 1: What is the connection between awards and allegations? That is to say, when allegations increase do awards increase or is there some other correlation?

Answer A:

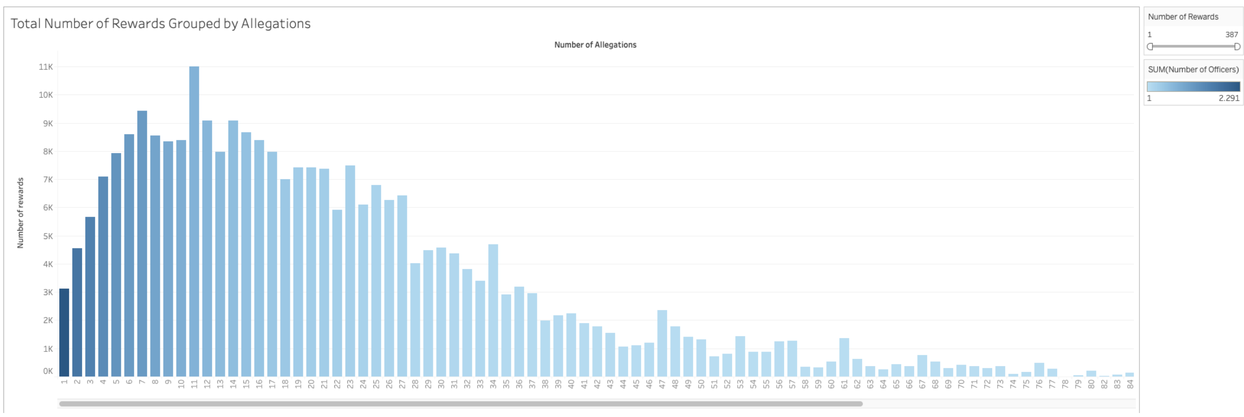
For the first part of question 1, we visually presented the relationship between police officer awards and the number of allegations they receive. We can conclude that there is a solid relationship between the number of allegations and awards that the officers receive. Additionally, we can identify that the data visualizations do have particular outliers that could potentially indicate “bad cops” (those with high allegation numbers and low rewards numbers). For this visualization and subsequent ones, we used some of our queries from Checkpoint 1. For each officer, we obtained a count of both their awards and allegations. The next two graphs show the number of awards per each count of allegations. That is to say for officers that have 17 allegations, the most probable number of awards was around 50. Most probable here is effectively the mode or the most common value of the dataset.

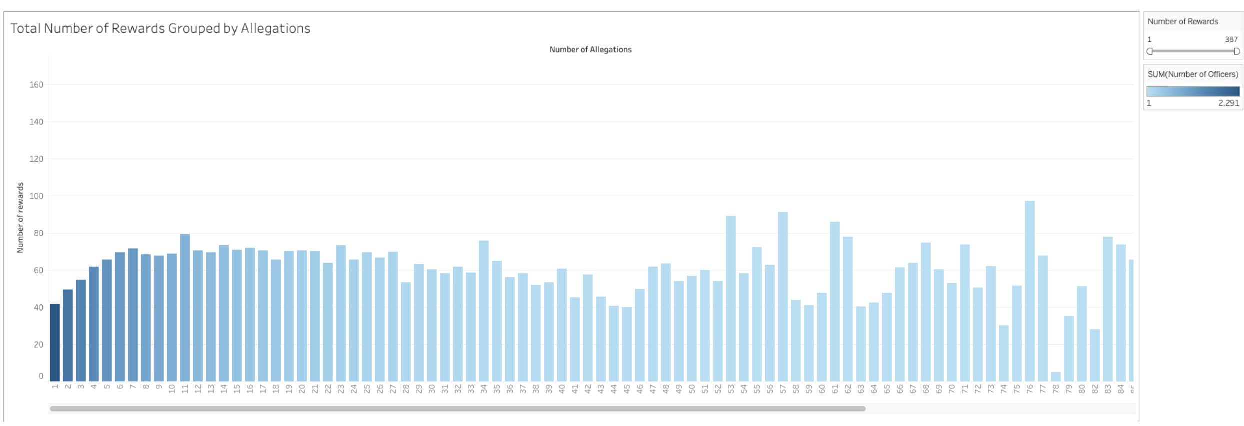

Number of Rewards Per Allegation
Number of Allegations
Number of Rewards
280
260
240
220
200
180
160
140
120
100
80
60
40
20
0
17 20 21 22 24 25 26 28 31 32 34 35 36 37 38 40 41 42 43 44 45 49 50 51 52 53 54 63 64 67 73 90
SUM(reward_number)
18
284


Number of Rewards Received Compared To Allegations Made Against Police Officers
20
17
21
22
34
49
43
35
36
50
63
32
64
45
24
37
56
51
90
73
55
44
25
39
54
40
52
67
47
46
53
42
41
38
31
30
27
28
26

Answer B:

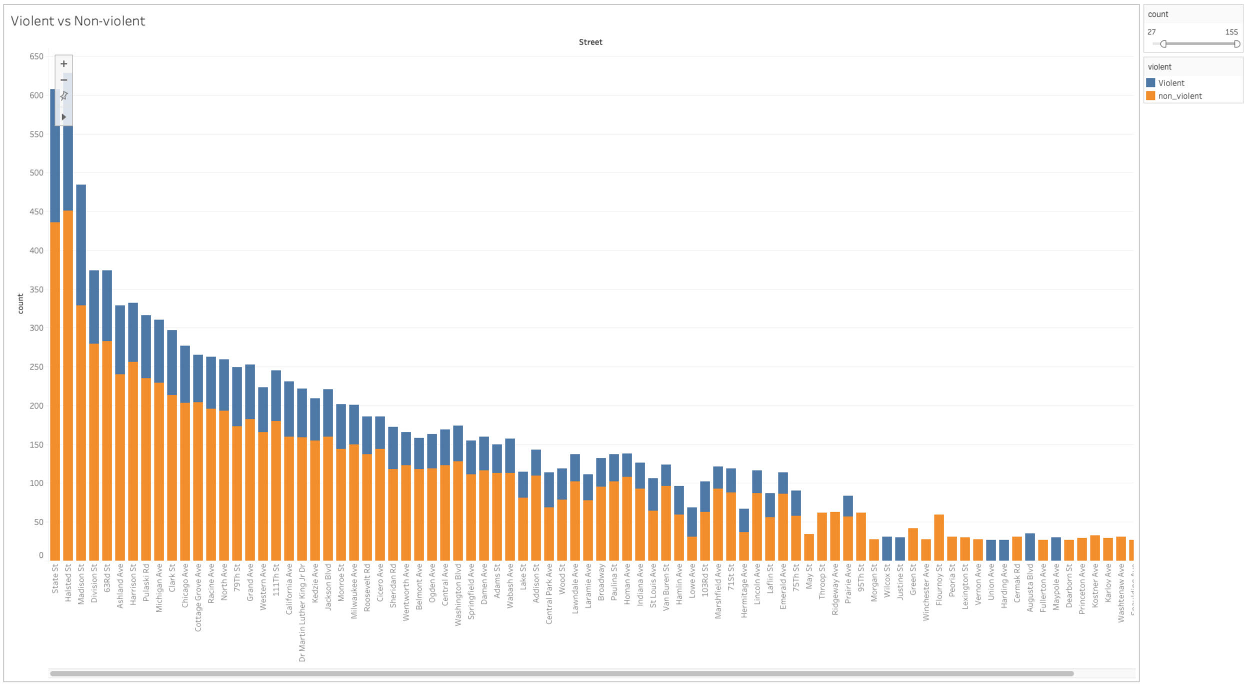
For the second part of question 1, we are generating two different graphs. The first graph represents the distribution of police officers' counts with a certain number of rewards and allegations. For example, we can clearly identify that when the reward number for each officer equals 1, their filed allegation number is also equal to 1. Furthermore, the count of police officers decreases when the number of allegations increases. Thus, we can understand the particular distributions given a certain number of allegations and rewards. For the second graph, we presented the average number of police officers rewards grouped by the number of allegations they receive. We can clearly identify patterns in the following figures where with the increase of allegations the average number of rewards increases significantly and vice versa.



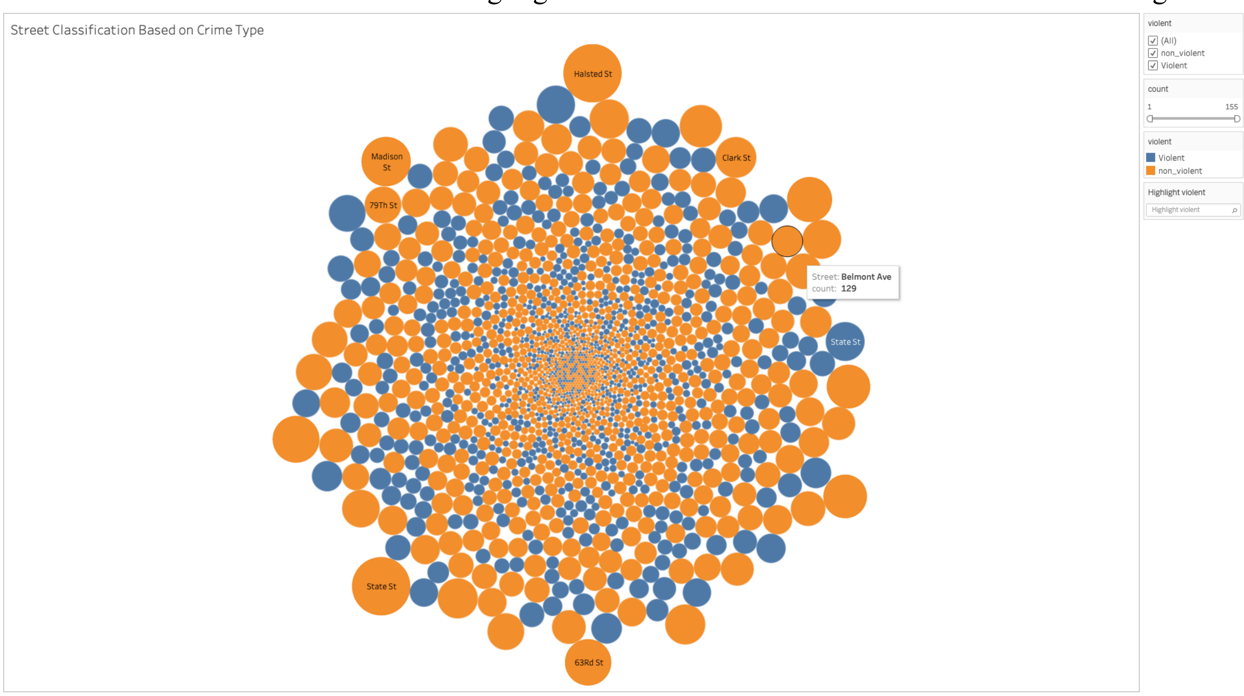


Question 2: Based on TRRs, what streets experienced the greatest number of incidents? Of these, in what percentage of these did the police respond violently (eg. the action response category was greater than or equal to 4)?

For these visualizations, we obtained the number of incidents per location (in this case, locations were all streets). Afterwards we determined based on the action category rating, whether or not each incident involved a violent police response (eg. a weapon was discharged, melee force was used, etc.).

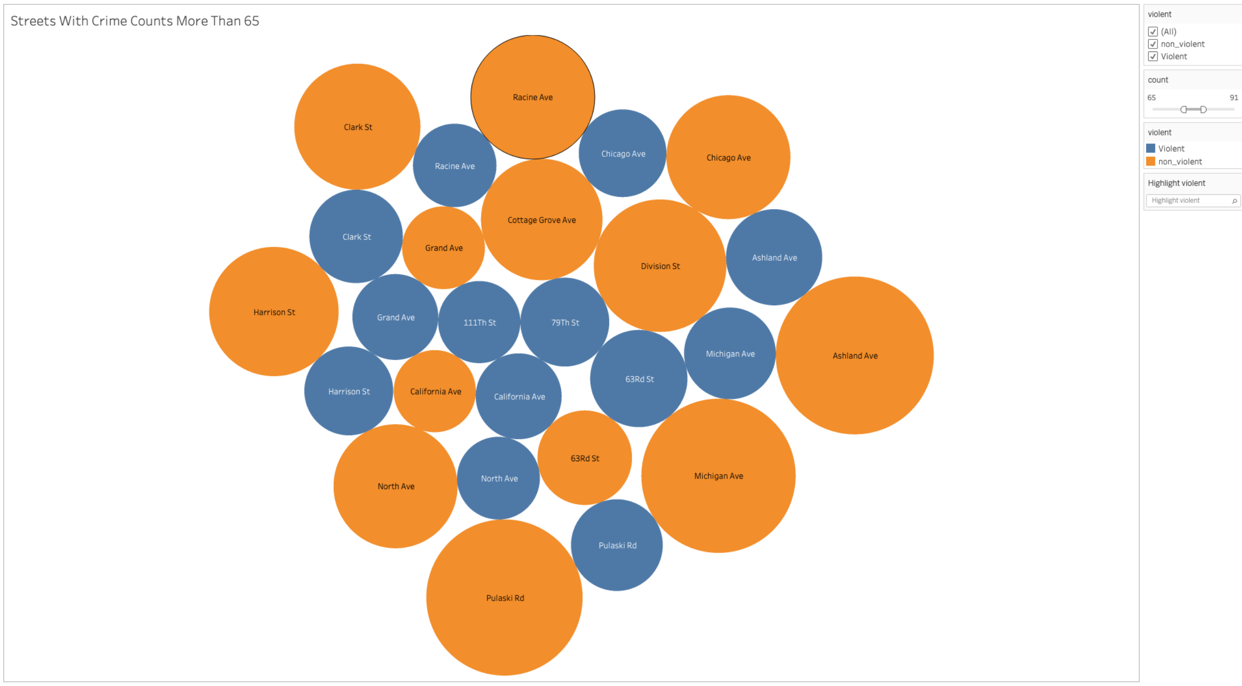


The above graph displays the number of incidents per street. It splits the incidents into violent and non-violent. The violent crimes are highlighted with blue and the non violent crimes with orange.

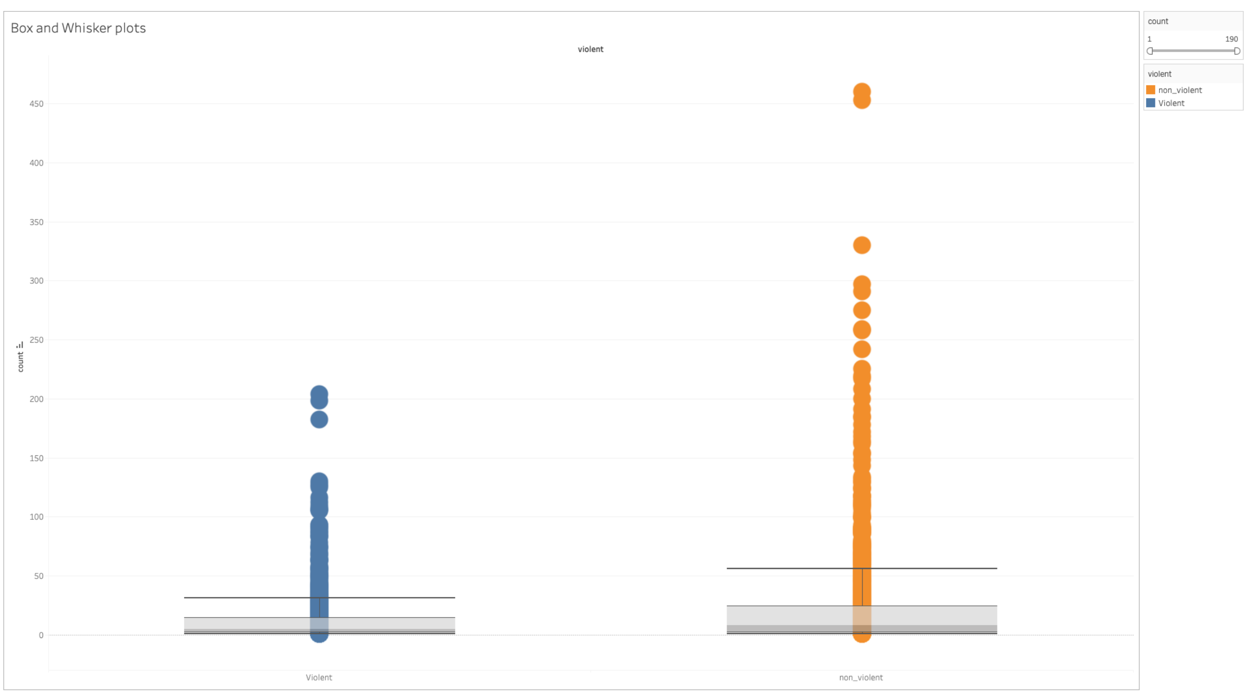


This bubble chart represents how dangerous several streets in Chicago are, based on the severity of the Tactical Response Report types. As a result, each street is represented by a single bubble whose size and color represent dangerous and violent or non violent incidents that occur in these streets. As a result, the size of the bubble is associated with the number of incidents and it's color with their severity.

For example, from the chart we can clearly identify that State Street is an area where a lot of incidents occur, both violent and non violent. So it would be best to stay on alert while near State Street.



This bubble chart follows the same principle as its predecessor. As a result, this chart displays the streets with incident counts (split into violent and non-violent for police response) above 65. The size of each bubble is the relative count. This shows, as expected, that non-violent responses are more common and also that many areas with violent responses also have a great many non-violent responses.

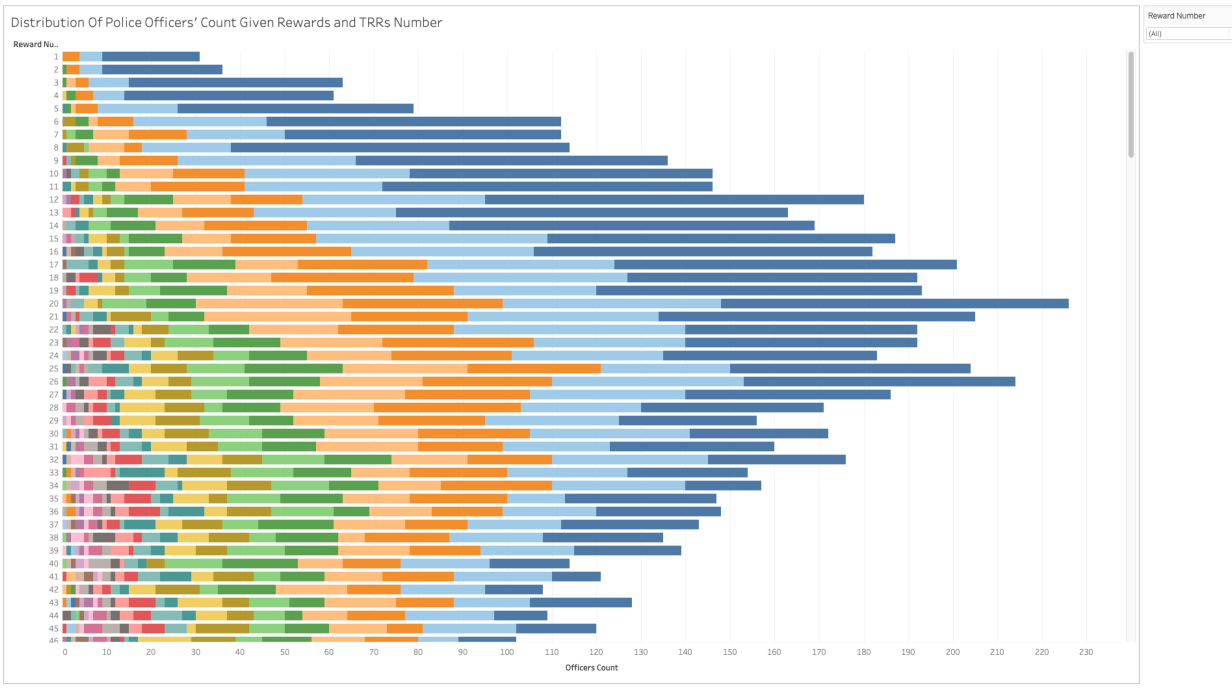


This box and whisker plot is mainly designed to show outliers. It shows that there are roughly 3 streets that fall outside the standard distribution for both violent and non-violent responses, and (although it's not displayed here due to the limitations of static images) the streets for both response types are the same. They are Madison Street, Halsted Street, and State Street.

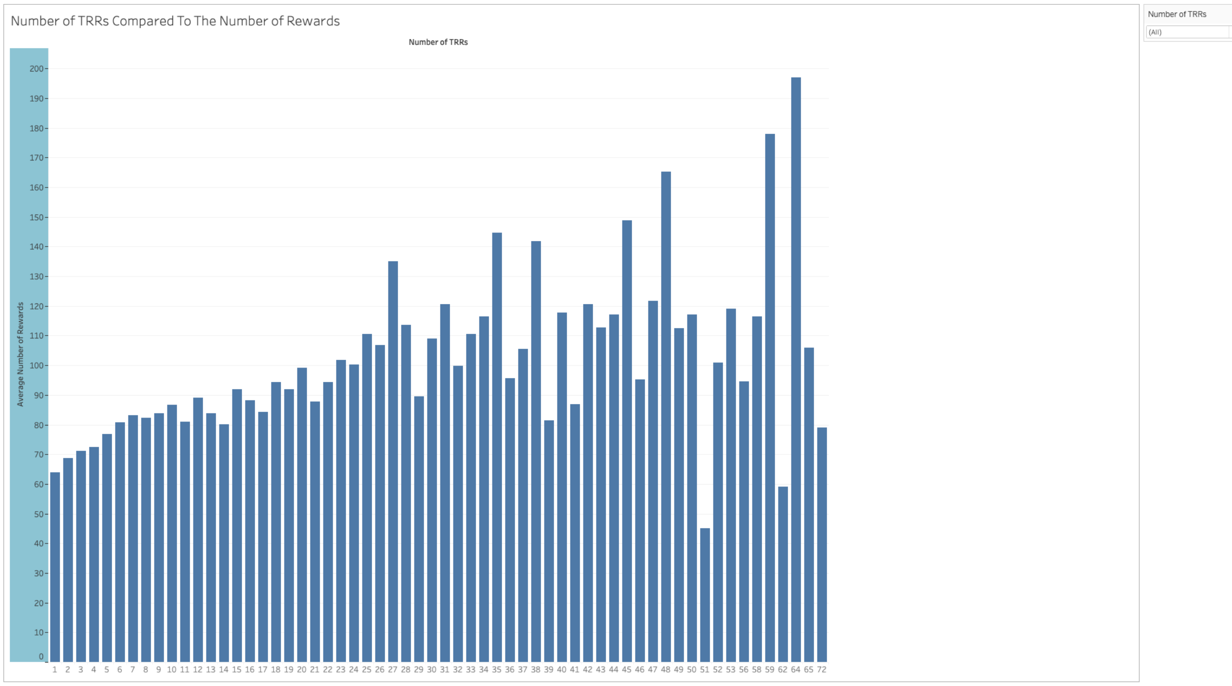
Overall, the visualizations give us a fair amount of insight into the police response per street. It allowed us to determine the places where the police responded violently most often as well as outliers of the normal distribution. Do these areas experience more crime or do the police simply react differently while in these areas? Perhaps this is a chicken and the egg situation...

Question 3: Is there a correlation between the number of tactical response reports filed by each officer and the number of awards that the officer received?

For question 3, we have successfully managed to generate two different charts in order to identify whether there exists a relationship between the TRR's and the number of awards that the police officers receive. In order to achieve that, we mapped the TRR police officer ID and the reward number that they received. In addition, we aggregated the police officer ID as to get the officer's count while grouping by Trrr-s and the number of rewards.



The above graph was generated by plotting officer count against the number of awards each officer earned. The data was then broken down into the number of TRRs they filed. The outermost color (dark blue) represents the number of officers who were involved in one TRR incident, the next color (light blue) is for two, etc.



This graph was generated by plotting the number of TRRs filed against the average number of awards per officer.

The two visualizations demonstrate something of a trend: as the officer is involved in more TRR incidents, they also will likely get more awards. One possible question would therefore be this: do officers who file more TRRs get more awards in recognition of their service in more dangerous situations or is it simply that officers will agglomerate awards and TRRs as their career continues (to explain: older officers have been involved in more events and thus have both more awards and file more TRRs).

Conclusion:

These visualizations gave a fair amount of new insight into the CPDP database. We have identified both a correlation between allegations and awards as well as one between filed TRRs and awards. Future extensions of this would likely need to decouple the time factor from the data. We would need to break down the data further by comparing officers of similar age groups. For example officers who have comparatively high awards and TRRs filed for their age group demonstrate, at least to some degree, that being involved in violent incidents is predictive towards getting more awards. By filtering on age group or number of years as a cop, we can remove "the number of experiences" factors.

We have also pinpointed the streets where the police acted violently most often. It is difficult to theorize about why these locations receive the greatest ire. Further research would benefit from using a crime database which contains the number of crimes per each street. However, this still does not answer an interesting psychological question: are the police responding violently because of the trepidation they feel in the area or due to actual danger in the area? This would likely require extensive psychological and sociological data on both the area and officers involved in the incidents. That falls outside the scope of this class and potentially has some morality issues.

Experience with Tableau

Tableau, while obviously a powerful tool, has its limitations. It honestly wasn't the easiest thing to use and was confusing at times. For example, much of our work was visualizing our queries from Checkpoint 1. To do this, we mainly imported in previous SQL queries. However, it is impossible to have multiple different questions answered in a worksheet, as when you attempt to use multiple queries Tableau will try to create a "relationship" from them. If there is no data to create said relationship, it will fail. Also in an attempt to use Tableau's normal querying, we discovered that relationships are somehow different from joins, and that joins must occur one level below in the data graph. This all goes to show that Tableau does not provide the best user experience.

Another place we found Tableau lacking was in the possible graphs. For our second question we wanted to create a bubble chart for each street (after filtering out those with low incidence), but also split the bubble in terms of violent vs non-violent, much like a pie chart. Much to our dismay, however, this was not possible. Another graph type that seemed unavailable was a diverging bar chart (eg. one where two categories are graphed in opposite directions). While these are somewhat specific tasks, it was also disappointing that Tableau didn't offer easy alternatives.