

Checkpoint 2 Write Up Analysis: Data Visualization

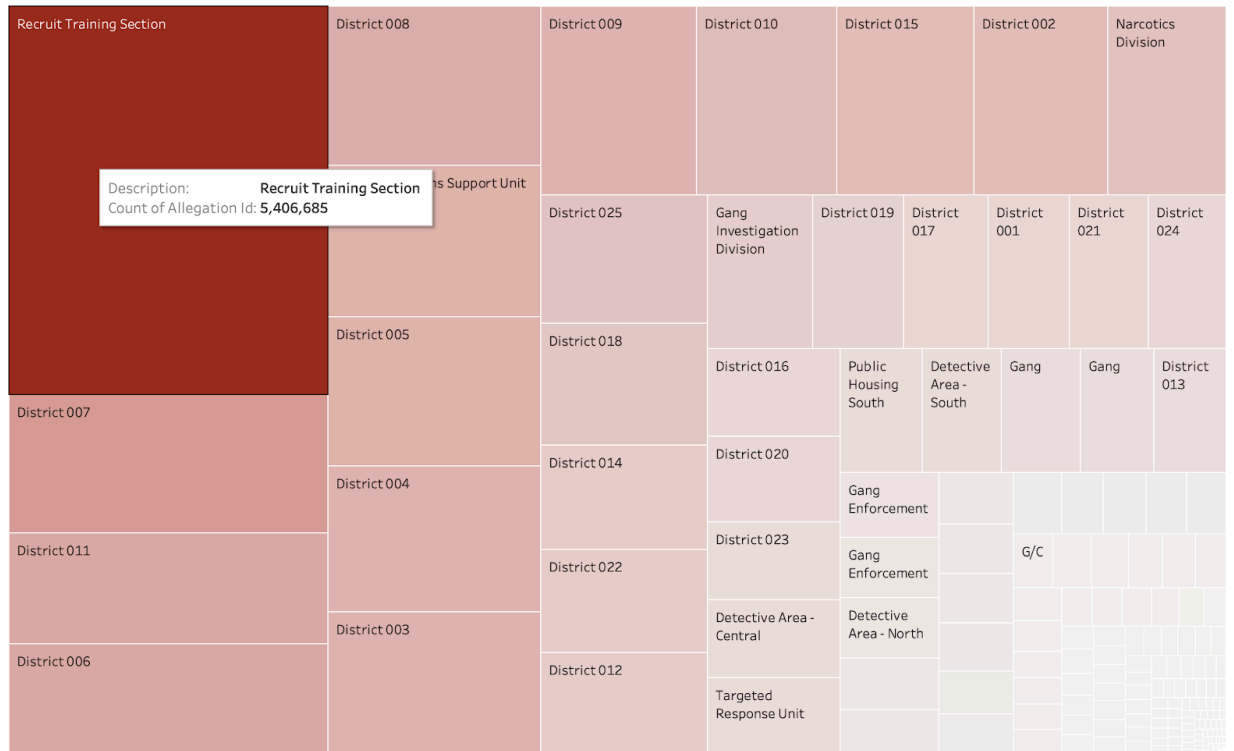
The theme of our project is to understand model example police departments, which is a pivot from our Checkpoint 1 theme. We want to better understand any potential systemic involvement in the misconduct of citizens by police officers and how the departments might or might not be turning a blind eye to these allegations. We can understand this by comparing across departments those that for example that are placed (patrol) in historically more volatile/aggressive neighborhoods with high crime rates. Perhaps one department in such an area has relatively low total complaints versus other departments in similar areas with higher volumes of complaints. Amongst other parameters, we will utilize department data on things such as: department and officer employment history, complaints against the department, promotions that happen in the department, employment of officers to departments as related to the district they are from (bias to hire certain officers from within their own district), and others as they become relevant.

For the checkpoint 2, we wanted to visualize relationship of police department unit with some factors that may contribute to discover new cases. Specifically, we focused on number of complaints per police unit, to find out whether it is distributed uniformly among the police units or is there any specific units which gets most of the complaints.

Working with Tableau was a bit challenging at first, as the interface was intimidating for first time users, but once we figured out how to do one thing, the rest just clicked. Linking tables and matching keys in Tableau is much easier for us than in SQL as everything is very visual and almost intuitive (for example, “inner joining” tables is much simpler in Tableau). The functionality of Tableau itself is quite nice too as much of the visual display portion is simply drag and drop. One main issue, however, was that we wanted to aggregate and visualize data across three dimensions, which was a bit challenging, as we only have our feet wet with Tableau and are not quite familiar with the advanced features. That being said, here are the questions answered via Tableau output:

1.

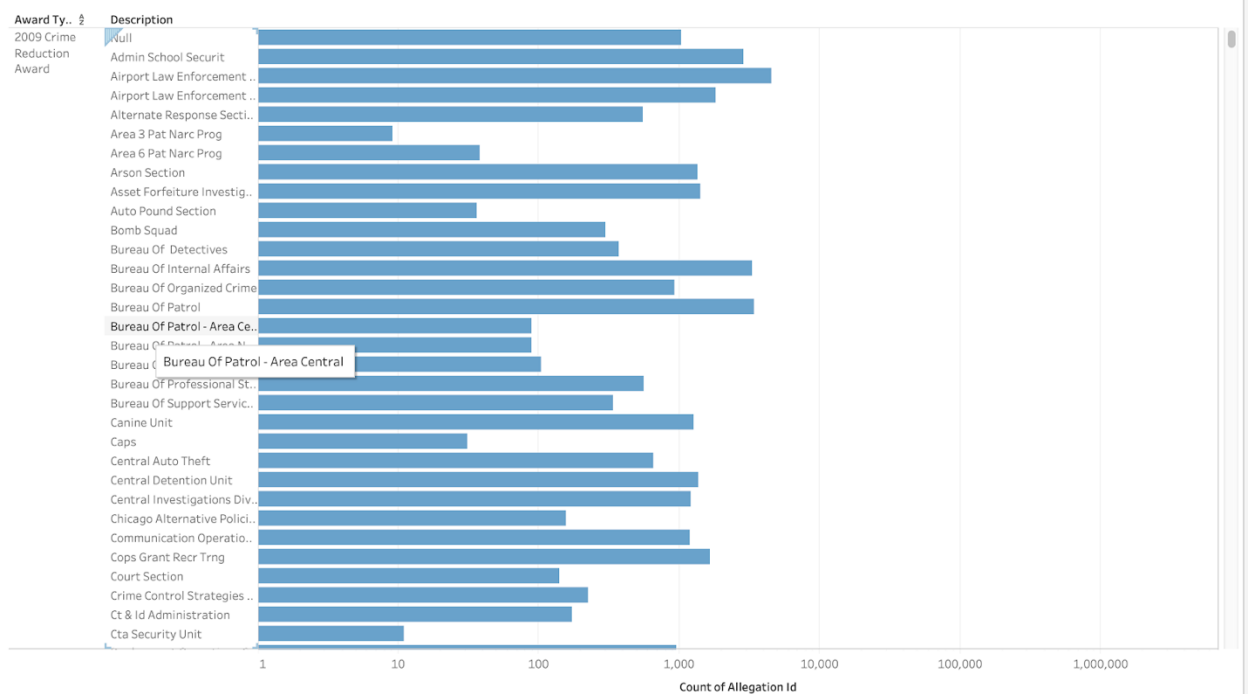
Number of Complaints per Unit



This is a tree-map of the total number of allegations per unit, with larger and darker squares representing larger shares of the data. Here, we see that the Recruit Training Unit has the most complaints out of all units, with District 007 and 011 being second and third. This intuitively makes sense, as recruits are more prone to error. Further investigation into this data also will help us identify anomalies across the relationships between units and misconduct.

2.

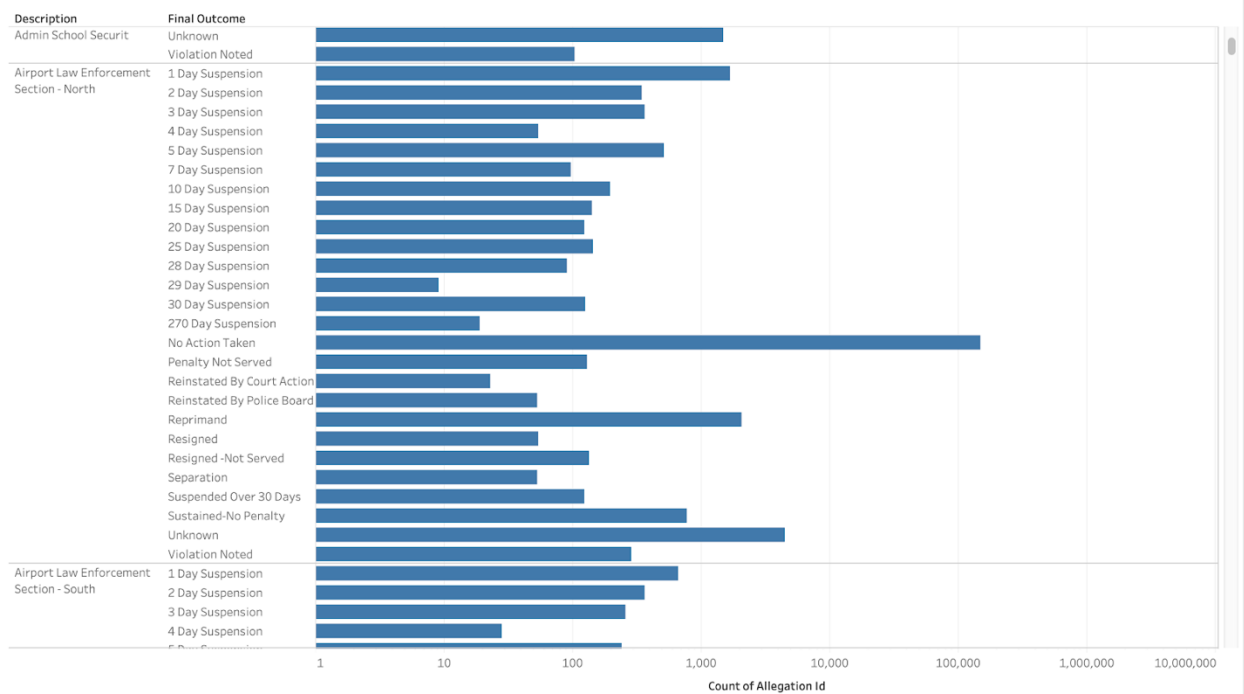
Total Number of Awards for Complaints per Unit



Our second question aims to answer the number of officers per unit that have received awards for corresponding complaints. This is to examine how the Police Department as a whole operates on both a departmental and per unit basis. This is interesting as the count of the number of allegations are represented on the bars, with each corresponding unit and award type (as you scroll down, you find segments for each award). Thus, the count of allegation id's represents the count of award "x" received per allegation.

3.

Outcomes Based on Number of Complaints per Unit



This is a bar chart depicting the final outcome of each complaint by unit. The Unit is presented in the first column, and the outcome on the second, with the count of allegations corresponding to each outcome. This shows us if and how officers per unit are disciplined on a per complaint basis. A deeper analysis into this (and next step) would be to get the total count of outcomes per unit to see if some units receive more discipline than others, and on that note, which units and why?

The question our group is seeking to answer using visual aid from D3 interactive visualization is: how many settlements did the police departments have to make annually to victims as related to allegations against the department? The D3 visualization we made shows two lines graphs to help answer our question. On the left-hand side, you can see we have incident count from 3500 to 7500. The blue line is associated with such dimensions as aided by the legend at the bottom as indicated 'Allegations'. On the right-hand side, you can see Settlement Count by \$ amount from 0 to 50. This is measured with the orange line graph as indicated 'Settlements'. This is an easy way to show allegations and total dollar settlement amounts per year from 2005 to 2015. There is not an apparent correlation between allegations and settlements. Allegations as shown are actually steadily decreasing over the 10-year span. Settlement counts seem to go in several year bursts of high activity, then followed by an equal amount of years with low settlement dollar counts. Note that this data is also not representative of the entire dataset as the Observable notebook is not able to withstand all of the data from the CPDB database in its entirety. We used 1,000 samples to illustrate this graph so one should be careful when extrapolating.

4.



Checkpoint 2: D3 Visualization

