Harkirat Gill, Susan Mulhearn
Data Science Seminar
November 4, 2020

Checkpoint 3

Zoomable Sunburst

In this analysis we continue to look at the distribution of allegation categories across units, beats, and crews. We first join officer allegations to the associative tables for unit, beat, and crew and join on the allegation category table. From this, we take the grouped units, beats, and crews and turn them into columns to get the counts of each allegation category from that group. These are the same generated tables from checkpoint 1. From this we create a script to convert the csvs into json to make generating the visuals easier. The aim here is to see in which groups we see a concentration of allegation types which can support correlated behavior as we look at smaller groups of officers like we see with crews.

When looking at the categories per unit, we see a relatively similar distribution across groups. The top three categories being personnel violations, use of force, and illegal search all having about the same proportion. With each unit capturing a large number of officers, a roughly similar distribution is expected. For all units these three take up about 70% of all allegations. There are about 8 allegation types occurring less frequently taking about 25% with some variation between units in the order. When we look at the level of beats, we sample 50 so the visual is not too cluttered. Again we see that the distribution of allegation types follows roughly how it was per unit.

Now when we look at crews, we see variations in concentration as expected. Again we sample 50 crews to make the visual readable. For larger crews, we see illegal search and use of force having a much larger percent of total allegations. When we look at smaller crews, the concentration starts to resemble the average with personnel violations being the most common. Of course, we expect this. What we aim to find is for smaller groups, we find a similar concentration to the one with crew, but to a lesser degree. This would suggest a similar phenomenon as in crews when people work together that may point to peer influences on an officer. Next we will group by beat and watch for officers working on the same day and look at the distribution of those groups to get better insight into how working together changes the allegation type distribution.

Network diagram

In this analysis we aim to gain evidence for potential peer influences on an officers' level of misconduct by looking at a network visual of a sample of officers in a unit. We get officer shifts for a particular unit (unit 8 in this case) and 30 officers. We then query all the shifts for those officers and join each shift with a temporary table mapping officer ids to the number of allegations they had per year. We do not use allegations per year here as the assignment and attendance table from which we query only has results from a 4 year period and both measures of misconduct level should tell us the same thing. With this, we create a python script where we group officers where their beat, watch, and shift day matches and count the number of times they match between two particular officers. We use this to generate a json file which we then use to create a network diagram with each node as an officer, the color representing the number of allegations and edges in the graph being strength of connection between two officers. We create one network visualization and see a relatively sparse connection with the highest connected officers having a roughly average level of misconduct in the group. Because we sampled only 30 to make things run faster, there may be many connections not shown and so we will generate the same diagram for more samples across different units as well as adjust our query to capture all the connections for a small group of officers rather than sample 30 officers leading to the sparse and incomplete graph.