CS496 CP3 Findings

Visualization 1: We will visualize the 25-50 (undecided on number, will decide based on how cluttered the plot will look) most accused officers in a scatter plot. The X axis will show the number of allegations, and the Y axis will show the number of sustained allegations. On each dot, a hover will produce the specifics of the officer in a tooltip popup, displaying information such as but not limited to name, race, age, years on the force, district/beat assigned, and salary. We are interested in this for the greater project of exploring the general demographics of the largest offenders, and perhaps this can inform us which features are consistent across these most accused officers.

We settled on the low end of 25 most accused instead of the high end of 50 after exploring both and realizing that they show similar trends and anything higher will make the visualization too clustered. We limited our queries to active officers only because that makes the most sense for current findings. We have also excluded district/beat assigned data from our final visualization due to the sheer number of entries from the attendance data(over 18 millions) and the inconsistency it showed when it comes to which beats best accurately represent each police officer's assignment as there are usually multiple. All other queries/features are retained.

We hypothesized that on average, more allegations would lead to more sustained allegations, although the ratio between the two will be low overall since only about 5% of the allegations are actually sustained. We have reason to believe that the most frequent

offenders will be white males that are relatively older, with longer years of service but lower salary.

Our findings suggest that we hit the mark on the majority of our hypotheses. The scatterplot shows an overall positive slope where more allegations lead to more sustained allegations, with some major outliers above the slope(that is, more sustained allegation per allegation than average). 100% of our most frequent offenders are Males when it comes to both top 25 and top 50, with the most frequent Female offender having 65 allegation counts, less than half of the most frequent male offender. This ratio is a lot higher than the ratio of total Males in the Chicago active police force of about 76%. 14 out of the top 25 are White, a ratio of 76%, compared to a ratio of 49% of total White police officers in the force. Combining both statistics we just discussed, there are 14 white males in our top 25 officers with the most allegations, a ratio of 76%, compared to a ratio of just about 39% for the entirety of the Chicago police force. In our last checkpoint, we found out the average age of all active police officers is about 46, whereas the average age of our most offending officers is about 56, quite a bit higher. The average years of service of all active Chicago police officers is about 18 years of service, compared to our query of 29 years, a similar jump compared to our findings on age. One thing that did surprise us a little bit was the average salary; the average salary of the top 25 most offending officers is $105291, compared to an average active police officer salary of $90180, about 10 grand higher. However, this does seem consistent with our age and years of service findings, where older age correlates to longer years of service correlates to higher salary.

Visualization 2: We can look at the beats data for police officers on a choropleth map, and when hovering over a particular beat, we can inspect a more detailed breakdown for the police officers in that beat when it comes to certain factors we deem relevant for our predictions such as TRR to CR ratio, complaints to award ratio, average training hours, and salary changes over time.

Originally, the above was what we were attempting to visualize for CP3's second visualization. We've revised it so that it is more clear, and such that it fits the tables provided by CPDB. We were unable to find salary changes and average training changes, and instead included TRR, CR, Complaint, and Award data by aggregating them by Police Beat. We calculate average CR, Honorable Mention, Internal Allegation, and TRR percentiles of the officers found in each beat, and visualize them with a choropleth and a dropdown. Upon hovering, a title should show the name of the beat as well as the average percentile for the selected statistic. Do note that we were only able to fetch data for 190/277 police beats in Chicago, due to missing data in CPDB. Data processing to arrive at the averages per each beat was done in a Python notebook that is attached to the submission folder.

Our findings from these visualizations are fairly inconclusive, to be completely frank. Across the board, there are nearly no discernible patterns among the data that we can see. This is likely due to the fact that police officers will often patrol outside of the beats that they initially attend to. In fact, when looking at the average complaint allegations on this choropleth, we see nearly no correlation with the number of allegations per capita for each police district d3.js demo (which I also authored); so what

can this data tell us about police conduct in general? We see averages of nearly .75+ for nearly half of the beats, and a very very significant number of <.25 percentiles (high percentiles are implied to be bad, e.g. worst offenders are at the 100th percentile).

Looking at the honorable mentions percentile graph can maybe inform us a little bit. Nearly every district appears to contain an average of the highest percentile of officers for honorable mentions! Perhaps this, and the previous datapoint on civilian allegation percentiles, can show that *there are no good apples*; despite being given different demographics (that also have extremely varied crimes per capita) for each beat, the fact that officers are still racking up fairly uniform high allegation percentiles *as well as* being honorably mentioned for them shows that in general, officers across the board are doing a poor job and getting rewarded for them. Perhaps this also gives evidence that there is a lack of culpability among officers.

TRR allegation percentiles are also fairly uniform across the board, with a majority of beats having a >.75 TRR allegation percentile. TRR allegation percentiles, with TRR meaning tactical response reports, can give evidence to the amount of force used by the officers from each beat. If we look at the data table import, Observable shows the distribution of TRR, which seems to be normally distributed with a mean of 60. We can contrast this with the internal allegation percentile, where the mean is normally distributed and centered at 35. Keep in mind that these percentiles are across all officers, including those active and inactive, but our data only includes active officers. When average statistics skew higher than an average of 50 (all of our data points, except for

internal allegation), that can heavily imply that officers that are being retained are *worse than average* for all of these statistics.

Lastly, we'll look at internal allegations. From the choropleth as well as the distribution chart, we can see that it is very different from the other statistics in the sense that the average is nearly .25, and there are many low internal allegations across the beats. Perhaps this can also speak to retention — that internal allegation rates are far more impactful in officer retention than civilian complaints. This disparity is obvious, and perhaps can show a general culture of not making many internal allegations in the first place (officer culpability is at question once again).