

For my visualizations, I sought to create scatterplots that show the relationship between metrics describing the composition of police units and the associated rate of allegations and TRRs.

#### Simplifications and resultant subsets:

- There are many units, not all of which are citizen-facing. To combat this, I subset my data to only include police units that are associated with Districts. Specifically, I subset my data to include only units with a descriptions containing "District".
- Since the composition of units is fluid, LEOs change units from time to time, I calculated composition metrics on a yearly basis, and subset from only 2010-2015 for query simplicity (I have multiple 400+ line queries due to aggregation and sub-querying). Additionally, for each year, I only considered LEOs that were in the unit for the entire year, filtering on the officer history table for when the officer started/ended membership in specific units.

Note on entropy: I was specifically interested in calculating the entropy of gender and race within units. To do so, I utilized the following entropy equation:

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

where  $p_i$  refers to the percentage of the total population represented by class  $p$ . Lower entropy scores represent more homogenous groups, whereas higher entropy scores represent more heterogeneous groups.

#### Analysis for Question 1:

- When looking at allegations per officer per year by different units, we see a fairly flat distribution across units and years. I included this chart as a baseline for comparison for other metrics that we will be investigating. The one thing that stood out to me, was that year over year within the same unit, the rate of allegations per officer did not change very much, indicated by groupings of the same color near each other.
- Allegations per officer by average service years shows us that the groups comprised of LEOs that have served on the CPD for longer have a lower rate of allegations. This could be for many reasons, such as when officers gain experience (and rank) they change to roles that potentially deal with citizens less often, or it could simply be that officers with more experience have learned to act in ways that do not result in allegations.
- We see a similar relationship when looking at allegations by officer by average years in specified unit; more time in the unit correlates to a lower rate of complaints per officer. This metric is related to the previous, as LEOs in one specific unit for a long time will also have more service years. Additionally, this chart may show that with increased time in a specific unit, LEOs can gain familiarity with citizens/areas/beats and learn to act in ways that do not result in allegations.

- For gender entropy vs. allegations per officer, the scatterplot showed a slight positive trend, but nothing significant. As the ratio of men and women in a unit nears 1, we see a slight increase in allegation rate.
- The most interesting chart, in my opinion, is the race entropy vs. allegations per officer chart. This scatterplot showed a positive trend, indicating that as the race representation of a unit increases, so too does the rate of allegations against the LEOs of the unit.

One specific confounding variable in the scatterplots discussed above is that each unit represents a different area of Chicago. Each area of Chicago is made up of citizens that act in different ways and have a different baseline level of allegations per officer. One way to further investigate this data would be to look at the same units over time as the composition of the unit changes (I will be doing this for the interactive portion!).

#### Analysis for Question 2:

Note: the scatterplots for question two are identical to those in part one, but the y-axis represents the rate of TRRs per officer in each units.

- To be concise, the general shape/direction of each of the scatterplots is essentially the same as for part 1.

#### General Analysis

My original hypothesis was that more heterogeneous police units would have lower rates of allegations and TRRs (per LEO). However, by looking at the scatterplots in each dashboard we see that this is not the case, at least when explored across units. The next step in the analysis will be to look at the changes by year as the composition of units changes. I will look into this implementation for the interactive visualization checkpoint.