

Checkpoint 1: Findings

Team: The Wise Lobsters

Members: Amy Booker, Blaine Rothrock, Akash G V S

Theme

Identifying allegations early in a repeater officer's career and exploring patterns that lead to increasing allegations over time.

Assumptions

These are values that may be altered as we move through the research project, but for purposes of this checkpoint we have chosen the following values

- **Repeater:** officer with at least 10 allegations
- **Appointed date range for officer subset:** 01/01/2000 - 12/31/2007
- **Age Groups:** <25, 25-34, 35-44, 44-54, 55-65, >65

Question 1

How can officers be categorized to show allegation count over time (year/year)?

SQL File: q1.sql

Step one is to find a subset of officers that fit what we are looking for. We want officers that we can reliably look at from the start of their career. The dataset is complete starting in the year 2000 and has allegation data up until February 2018. With that, we limited the officer data to active officers starting on the force (appointed_date) between 01/01/2000 and 12/31/2007, this gives each officer in the subset at least 10 years of active services.

Officers are categorized overall by age and years on the force. The age groups are 25-34, 35-44, 45-54, 55-65. The subset used covers the years 2000-2007. When splitting the allegation count over time, years on force, the user is able to gather patterns about which age group of officers has spikes in the number of allegations. The largest number of allegations of officers by age occurs between age group 35 - 44 (allegations count: 19680).

Another split we include in the number of allegations year over year (years on the force). In aggregation, we look at averages over 10 years, which is the minimum number of years each officer in our subset has on the force. An interesting finding from this is there is a slight increase in average allegations over year two and three with the highest concentration in years 1-5.

Question 2:

What is the average number of allegations for different age groups of officers?

SQL File: q2_age.sql & q2_years_on_force.sql

In question 1 we created a view that organized officers by age group and years on the force from our subset of active officers with at least 10 years of active duty. For this question, we want to build on this subset and ask the question, what is the average number of allegations for each grouping. For our 10-year age grouping, we did not find much difference across age but did find a slightly larger average among officers in the 35-44 year age group (9.6 versus ~8.7 in the two older groups).

We also looked at allegation counts grouped by years on the force and, not surprisingly, found an increasing pattern ranging from 7.05 (10-11 years) to 11.62 (18-19 years).

Question 3:

What is the most common first allegation for officers overall?

SQL File: q3.sql

Our theme is focusing not only on repeaters but the road to repeated behavior. This question, again, will serve as a baseline for identifying where repeated officers start by pinpointing their first allegation. We did this by identifying the officer's first allegation by incident date, extracting the allegation category, then grouping by the category. Again, we limited this search to our subset of officer, but the query could be easily updated to include all or a different subset.

A large portion of allegations are categorized as Operational/Personnel Violations. There is a consistent surfacing of the data where, first allegation, repeaters allegations, and the most common allegation. The top three categories for all allegations and first allegation within our subset are as follows:

- Operation/Personnel Violations
- Use Of Force
- Illegal Search

Operation/Personnel Violations is a very broad allegation type, which could explain why it's the number one. For future research, it may be beneficial to focus on more meaningful allegation categories, like Use of Force, Illegal Search, False Arrest or Lockup Procedures. Or using other data sources to extract more meaning from the overarching category type.

Question 4:

What percentage of officer's first/early allegations are reported with repeaters?

SQL File: q43.sql

There are two variables that dictated the complexity of this query. The first is defining what a repeater is. For purposes of this exercise, we chose an arbitrary number of allegations, 10 total over the officer's career. We imagine as we continue to research, we will explore different values and new constraints for this value. The next is determining when an officer becomes a repeater for this we used the n th incident date where n is the number of allegations that makes a repeater.

This query was executed for the officer's first allegation only in our dataset but was compared against all officers. So for example, if an officer's first allegation was with an officer that does not fall into our dataset, it would be counted. We counted the number of first allegations from our officer subset that occurred with a repeater and in the allegation happened after the repeater's 10th allegation.

The result of this query is 18.62% of first allegations in our subset happened with a repeater. This is a significant amount and we would like to continue to explore the results of this. This query could be altered for various different approaches including refining the definition of a repeater or limiting to the type allegation.