



# The Wise Lobsters

Iram Naaz Khan, Blaine Rothrock, Akash G V S



# Theme

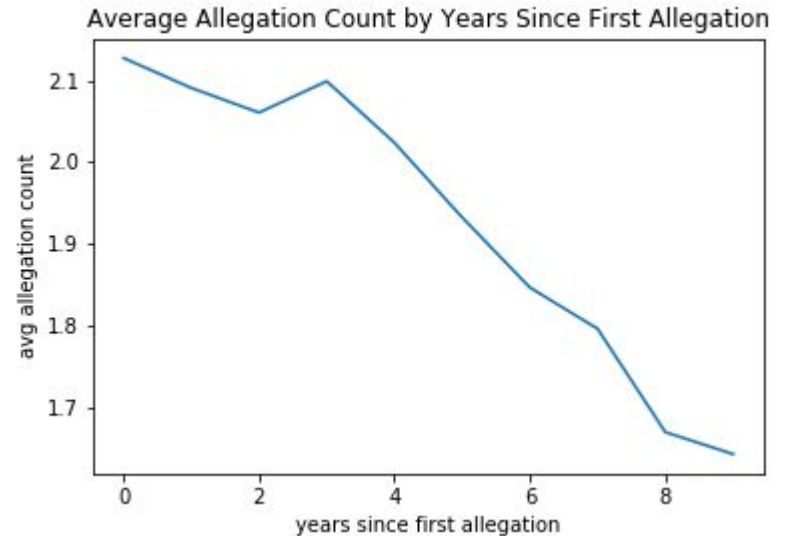
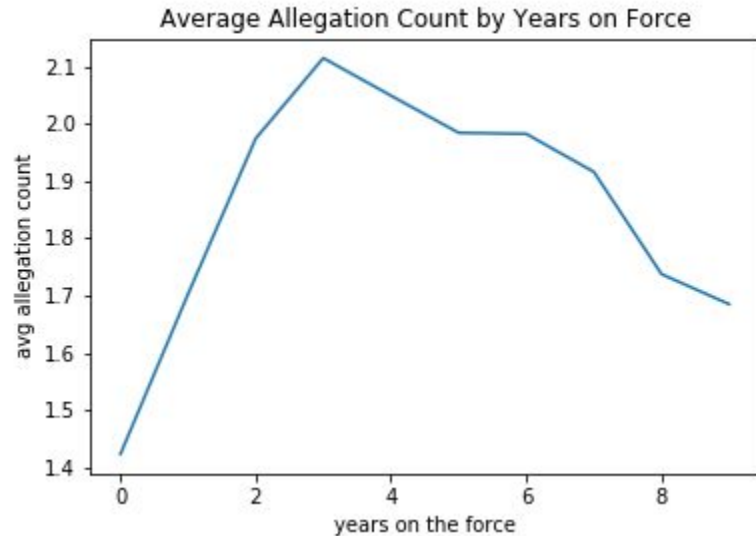
*Identifying allegations early in a repeater officer's career and exploring patterns that lead to increasing allegations over time.*



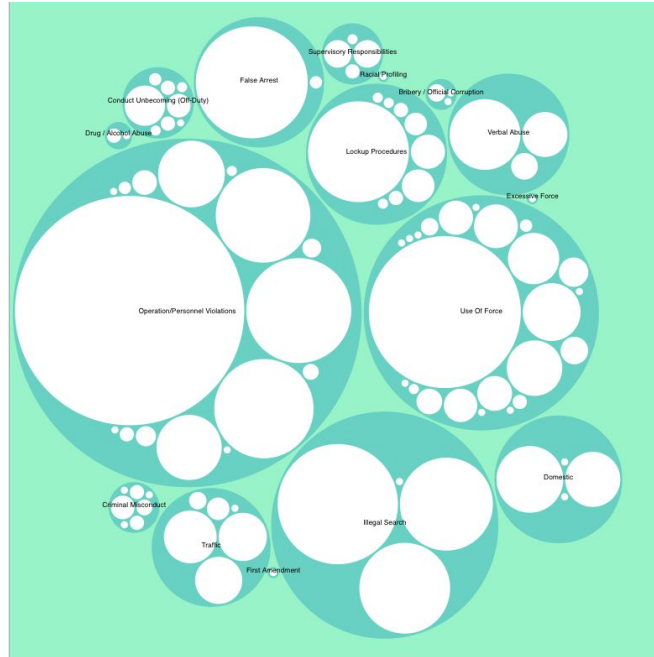
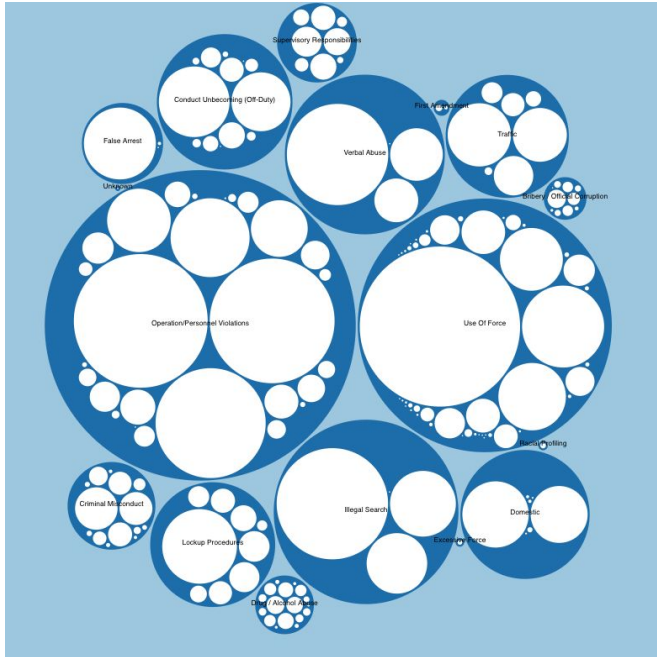
# Officers & Allegations

- Officers that have at least 10 years experience & full allegation data
  - Appointed between January 1st, 2000 and December 31st, 2007
  - Are active at the time of the last data upload (early 2018)
  - Years on the force is used for the time metric
- Officer's First Allegation
  - A main metric we use to measure an officers early career.
  - Civilian reported
  - Often we exclude common non-severe categories (Operation/Personnel Violations, Traffic, etc.)

# Observations: Allegation Count



# Observations: Allegation Categories



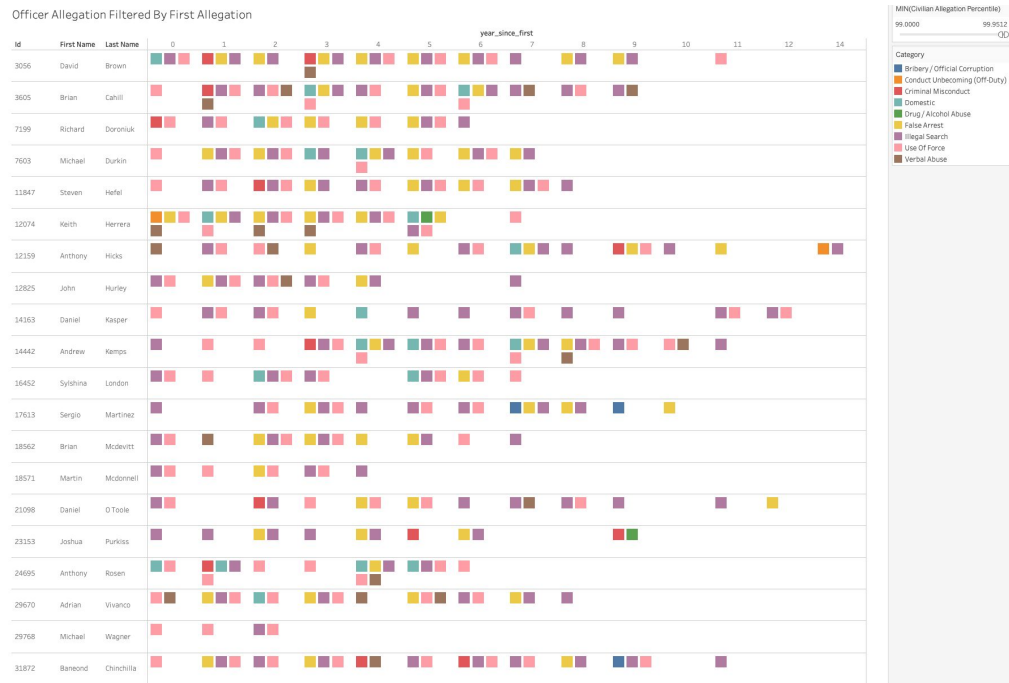
Left: first allegations  
Right: all allegations

# Analysis: Allegation Categories

What about the first allegation  
may leads to repeated behavior?

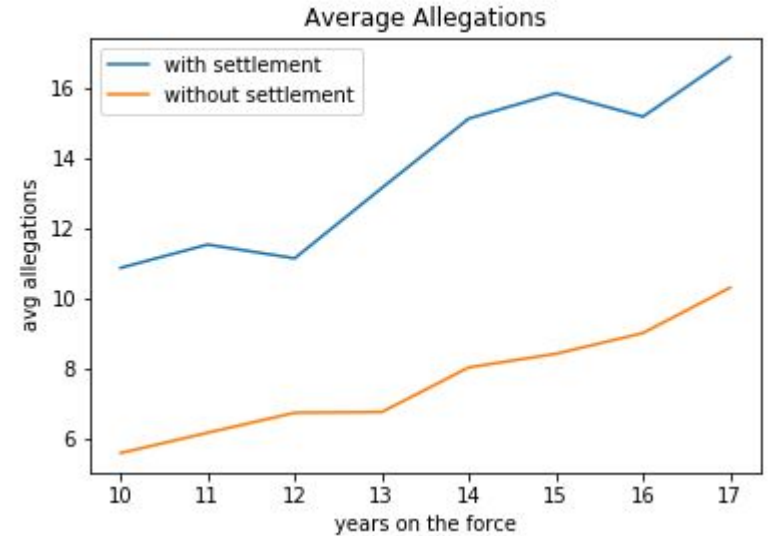
**Conclusion:** In generally officer's  
with violent first allegations tend  
to repeat throughout their career

On average, the first servere  
allegation category makes up **41%**  
of the officer's total allegations  
and **88%** of their total severe  
allegation



# Observations: Settlements

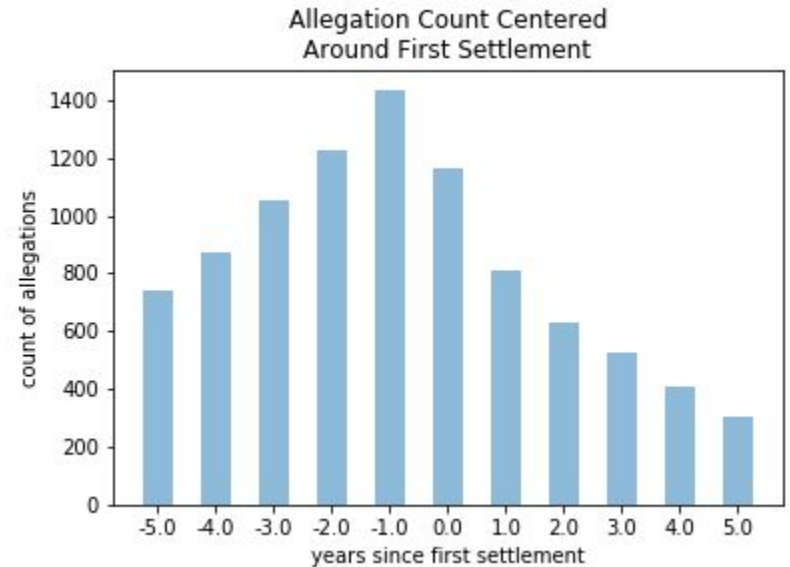
Officers with a settlement tend to have more allegations year over year.



# Analysis: Settlements

Count of allegations leading up to and after the first settlement. (officers within the subset)

**Conclusion:** Allegations counts tend to decrease after a settlement.

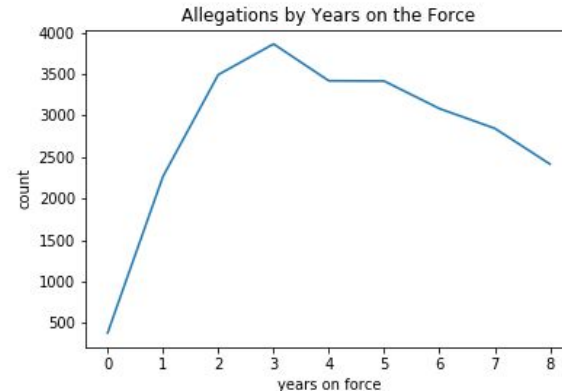
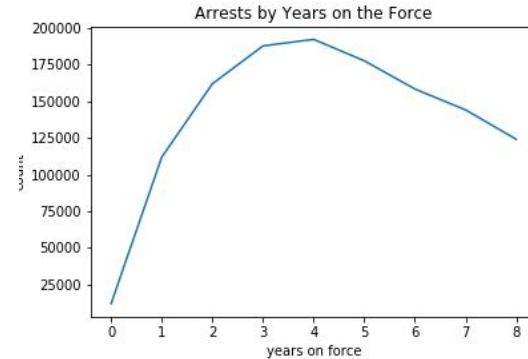




# Analysis: Arrests

Count for number of arrests against a particular officer over a year along with count the number of allegation for the arrest.

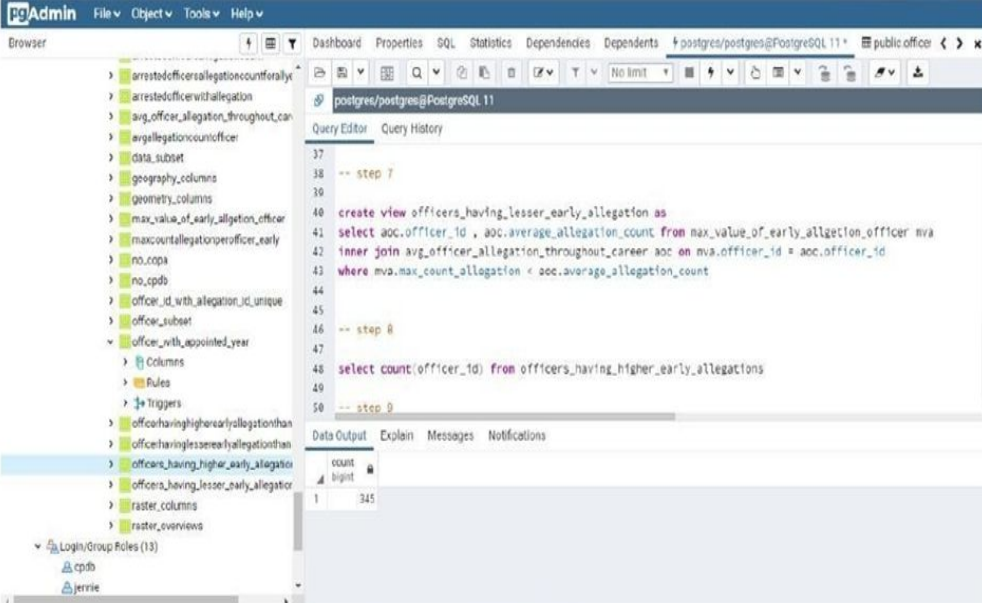
**Conclusion :** It has been observed that a large number of arrests have been done in between 3 to 6 years of serving on the force. It implies that for the officers with varied appointment dates and timelines of service, a large number of allegations have been done in this range.



# Analysis: Arrests

Finding out officer with high number of early allegation and later on figuring out whether such officers have higher average of allegations throughout the career.

**Conclusion :** A high number of allegations with an arrest early in an officer's career lead to a higher average of allegations



The screenshot shows the pgAdmin 4 interface. On the left, the 'Browser' pane displays a tree of database objects, including 'officers\_with\_appointed\_year' and 'officers\_having\_higher\_early\_allegations'. The 'Query Editor' pane on the right contains a SQL query with comments for steps 7, 8, and 9. The 'Data Output' pane at the bottom shows the results of the query, which is a single row with a count of 345.

```
37
38 -- step 7
39
40 create view officers_having_lesser_early_allegation as
41 select aoc.officer_id , aoc.average_allegation_count from max_value_of_early_allegation_officer mva
42 inner join avg_officer_allegation_throughout_career aoc on mva.officer_id = aoc.officer_id
43 where mva.max_count_allegation < aoc.average_allegation_count
44
45
46 -- step 8
47
48 select count(officer_id) from officers_having_higher_early_allegations
49
50 -- step 9
```

count
345

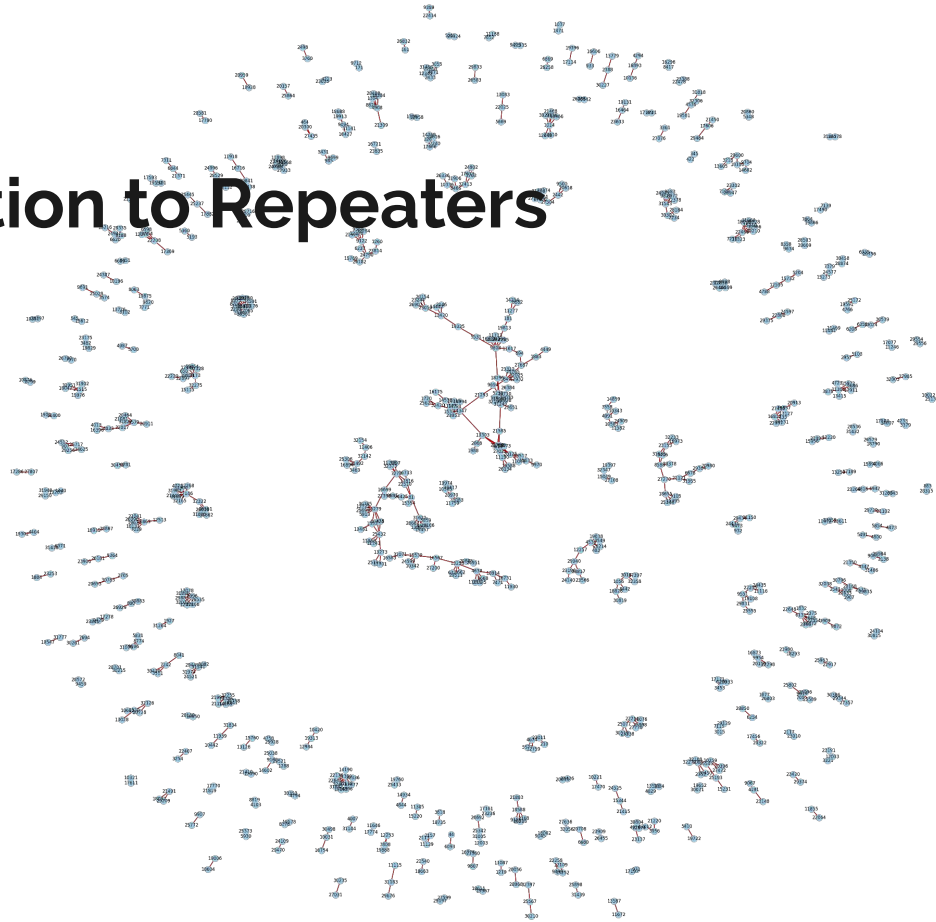


# Network of First Allegation to Repeaters

Is the a network between an officer's first allegation co-accused with repeater:

- **Nodes:** Officer (subset or repeater)
- **Edges:** Allegation co-accused with repeater
  - **In Degree:** repeater
  - **Out Degree:** subset officer

Identification of officers with the highest in-degree and work backwards to find clusters





# Network Findings

This analysis was able to answer the goal of finding the network and we were able to relate it to our theme of Identifying allegations early in a repeater officer's career and exploring patterns that lead to increasing allegations over time.

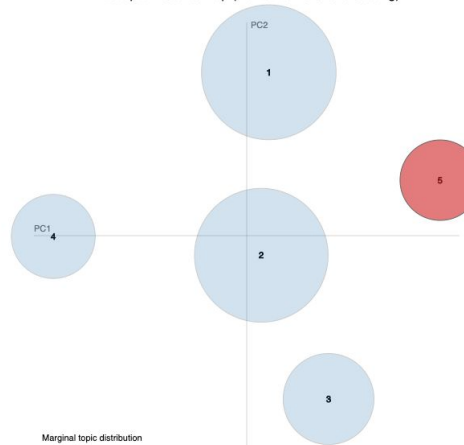
The formation of these networks can provide a point to our theme that an officer's first allegation, if with a repeater, can form a path towards repeated behavior. If the first time offending officer in the clusters are repeaters themselves and have allegation counts above the norm, then this could be an indicator early in an officer's career of how likely they are to have repeat behavior. The following are some questions we can explore in future analysis. We can look deeper into the cluster being formed and find the average number allegations for officers in these clusters and identify whether they are higher than average? We can identify are there types of allegations that are most likely in these clusters?

# Analysis: Use of Force Document Tagging

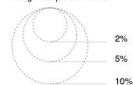
- Latent Dirichlet allocation model to categorize documents

id	term1	term2	term3	term4	term5	term6	term7	term8	term9	term10
0	0	rule	decision	follow	conduct	hear	write	therefore	hearing	suspension
1	1	vehicle	beat	car	stop	driver	drive	weapon	violation	traffic
2	2	sustain	member	disciplinary	employee	support	incident	witness	command	file
3	3	weapon	member	fire	discharge	battery	incident	injury	beat	force
4	4	go	sustain	incident	take	leave	work	back	witness	call

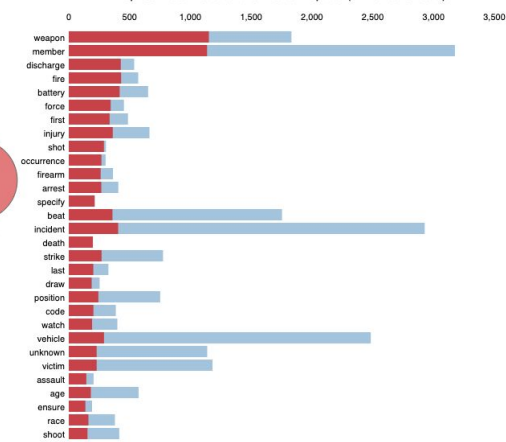
Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution



Top-30 Most Relevant Terms for Topic 5 (11.2% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) \* [sum\_i p(i | w) \* log(p(i | w)/p(i))] for topics i; see Chuang et. al (2012)  
 2. relevance(term w, i, topic l) = A \* p(w | l, i) + (1 - A) \* p(w | l, i/p(w)); see Sievert & Shirley (2014)



# Challenges

- Comprehending all aspects of the data
- Assumptions made on data before integration
  - Arrests
  - Document Tags
- Spark, Graph Analysis & Machine Learning are different learning curves
- Breadth over depth affects quality of conclusions (but we learned a lot!!)



## General Conclusions about the data

- An officer's first allegation type seems to be a pattern throughout their career
- Officers with early severe allegation tend to remain high risk throughout their career
- Repeaters are common and may influence officers early in their career



# Future Analysis

- Dig deeper into the co-accusal graph and explore clusters of repeaters
  - What percentage of officers co-accused are repeaters themselves
- More granular allegation category analysis based on allegation name
  - Cleaning and categorizing allegation names to get a larger set of categories
- Incorporating TRR reports of an officer's early career
  - We may do this to some extent for checkpoint 5 feature engineering



**End.**

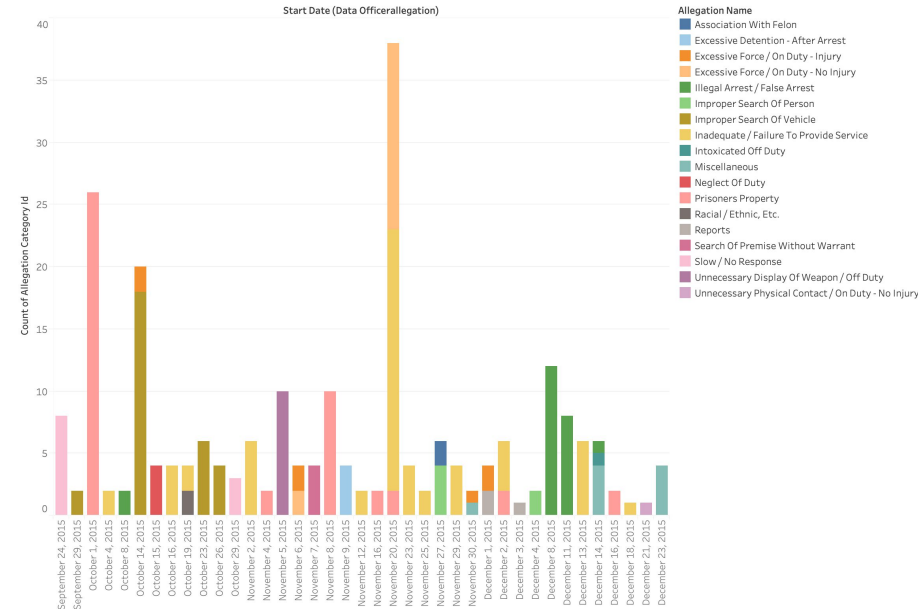
---

# Inconclusive Analysis: Media Events

Reported allegations around the release of the Laquan McDonald shooting video.

**Conclusion:** No drastic change in reported allegations. Interesting spike a few days before the video release.

Allegations By Category Around 11/24/2015



Count of Allegation Category / Id for each Start Date (Data Official Allegation) (MDY). Color shows details about Allegation Name. The data is filtered on Start Date (Data Official Allegation), which ranges from 9/24/2015 to 12/24/2015.

# Inconclusive Analysis: Allegations to Awards

Using graph analytics to link allegations to awards.

- **Nodes:** Allegations Categories, Award Types
- **Edges:** Award occurring 60 days after a allegation

**Conclusion:** It's not rare that awards are given after allegations, but categorizes and types are not descriptive enough to draw conclusions

