Final Project Presentation

Robert Loza, Jacob Wat, Bodhisatta Maiti

The Freedom Donkeys

Project Theme

“compare and contrast the common conditions and precursors for discipline in civilian vs officer reported complaint reports. “

CP1: SQL Analytics

Initial Questions

1. What is the number of officer disciplines for each allegation category?
2. Which categories have seen the largest percentage increase in disciplines as time has gone on?
3. How has the difference between the number of department and civilian complaints changed for different categories over time?
4. How have the number of civilian/police disciplines changed in relation to the number of allegations for specific “problem areas”?

Analysis

- Q1
  - From our executed answer, we can see that the categories with the largest number of disciplines are "neglect of duty", "miscellaneous", and "associated with felon".
  - miscellaneous is non-documented
- Q2
  - over a period of ten years (between 2000-2010).
  - "Inadequate / Failure to Provide Service" allegation has risen 5.54%, the most over the selected ten years.
  - "Slow / No Response" allegation has dropped by 2.77% during the same period.
- Q3
  - for most categories, the difference between civilian and officer complaints has seemingly decreased. However, this is due to a lack of recent data so it seems like there were complaints in the past. For example, the difference in “Use of Force” complaints dropped from 1174 in 2001 to 133 in 2017. “False arrest” complaints similarly dropped from 98 to 83 in the same time period
- Q4
  - With our current SQL code, we find the number of officer and civilian complaints for each beat. You can see based on the data that there is very little overlap between the beats that have the highest number of civilian complaints and the number of officer complaints.

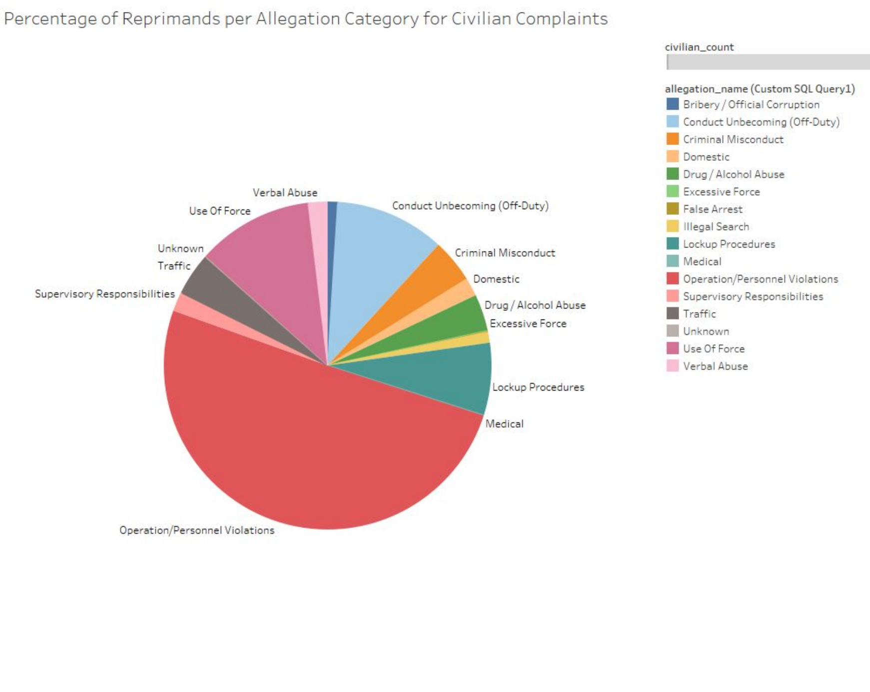
Beat ID	Number of Officer Complaints	Beat ID	Number of Civilian Complaints
160	820	2	1773
132	586	132	1532
261	464	261	1477
32	2	54	23
171	1	53	9
185	1	55	6

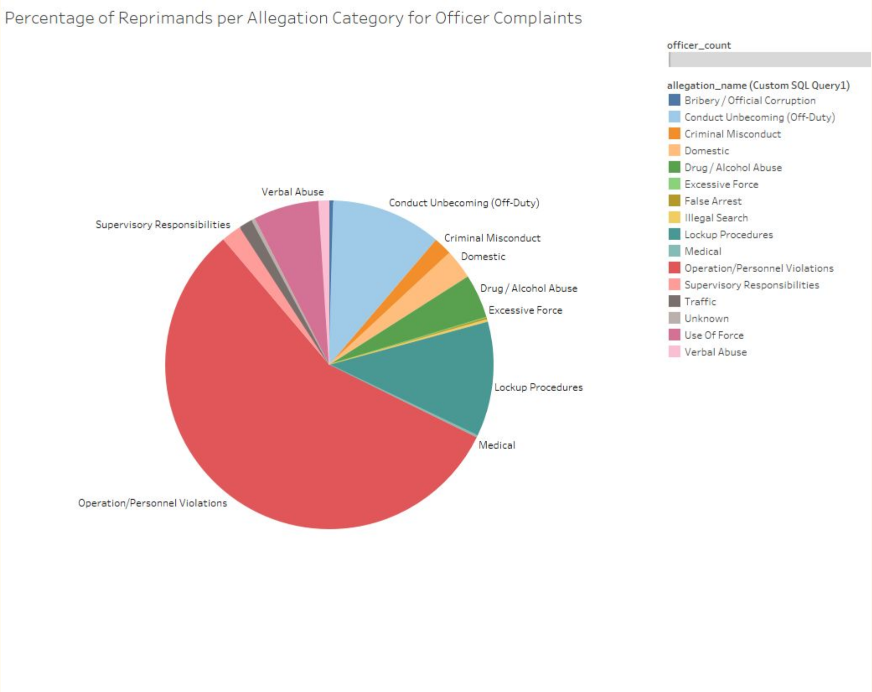
CP2: Data Exploration

Initial Questions

1. What is the relationship between the number of officers disciplined in each category of allegation for police/civilian reports? (2 Pie Charts - 1 for each type of report)
2. Use a symbol map to show the number of officer disciplines (per capita) from a civilian/police allegation for each district in Chicago? (2 symbol maps - 1 for civilian allegations and 1 for police allegations)

Data Visualizations

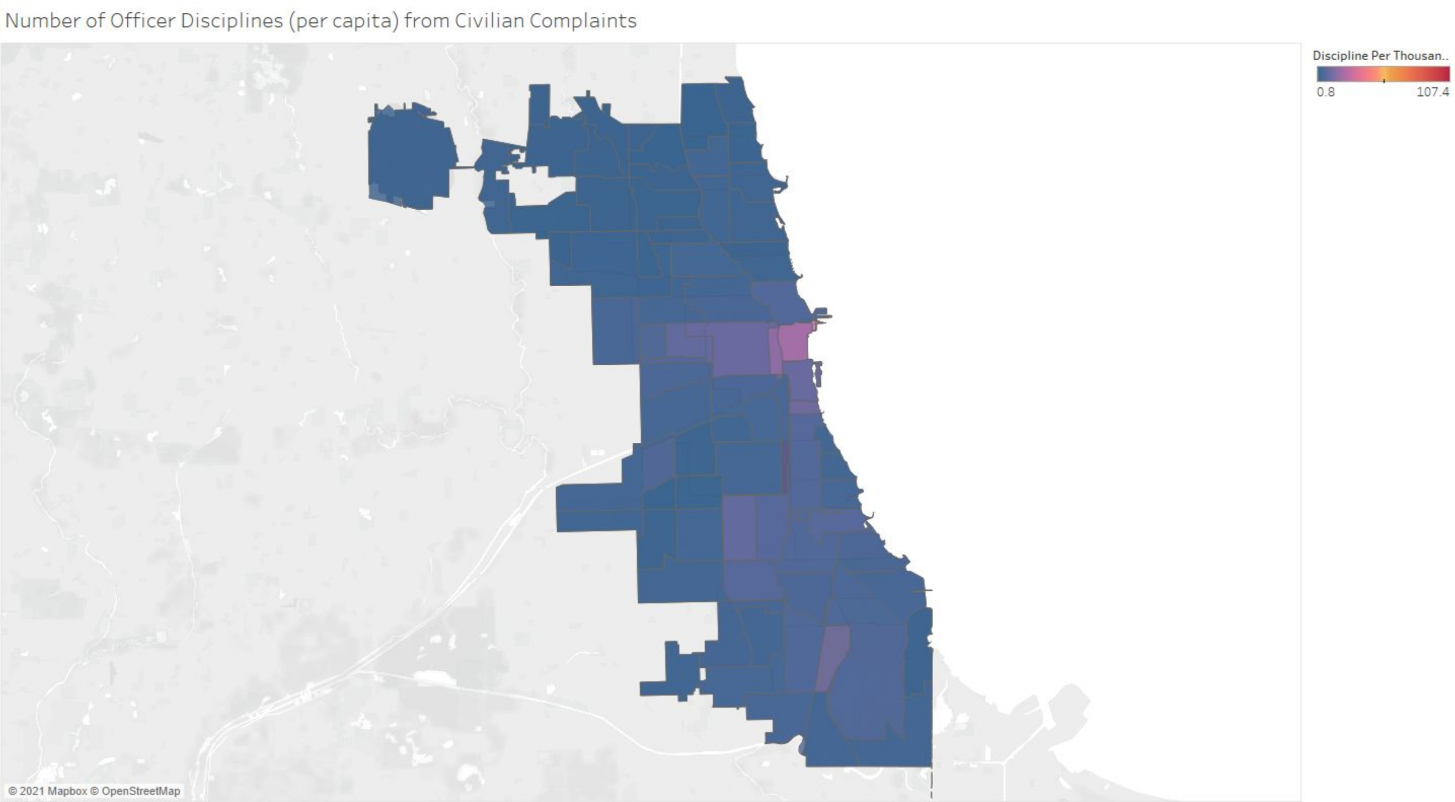




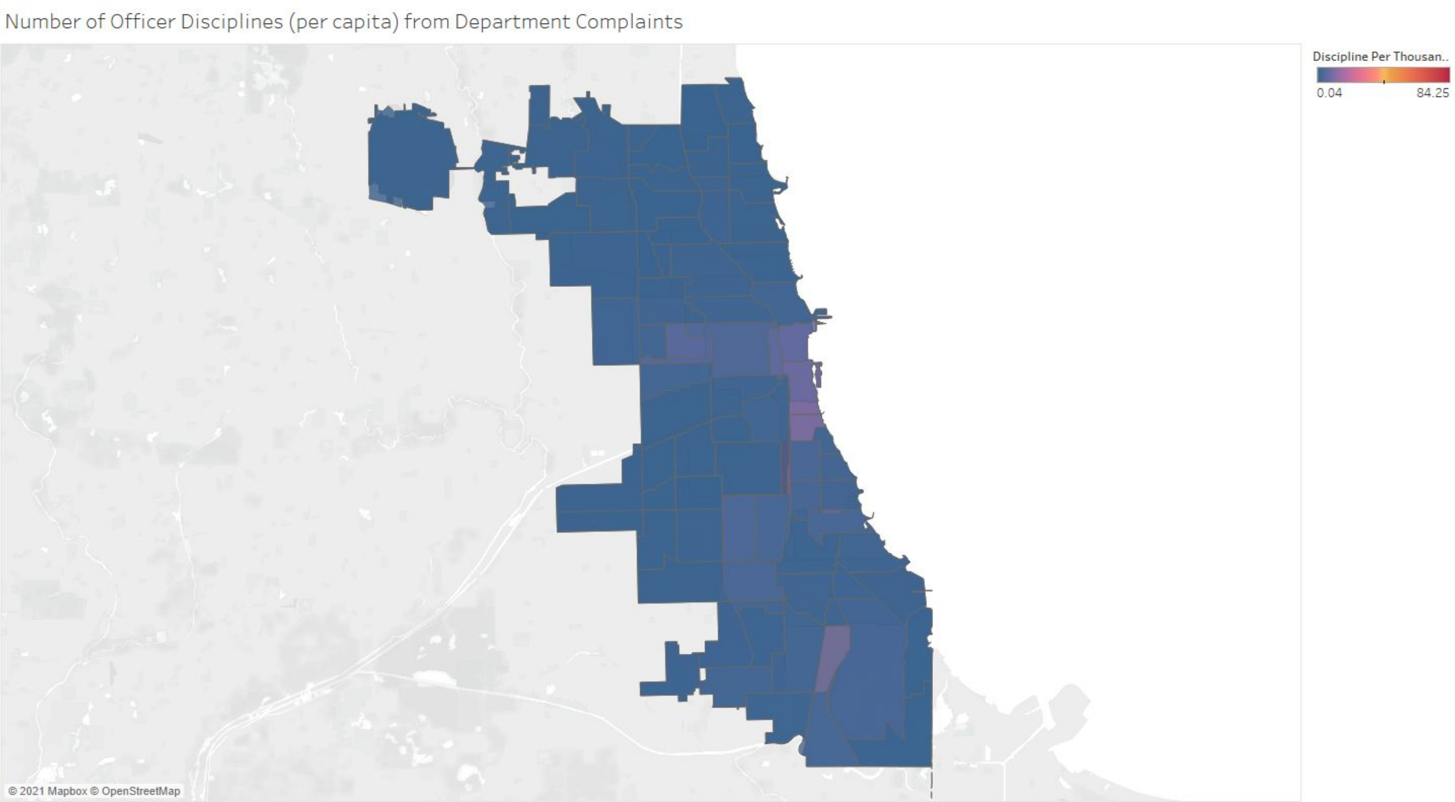
Analysis

- Several of the largest categories for both police and civilian complaints match up
    - The four largest categories approximately match up for both graphs
    - Operation/Personnel Violations, Use Of Force, Conduct Unbecoming, Lockup Procedures
- There are officers filing complaints who have experienced or have seen the same things civilians do
- However, since we are using the allegation category instead of the allegation name, we lose a lot of potential data or correlations there.

Data Visualizations (cont.)



Data Visualizations (cont.)



Analysis (cont.)

- For civilian allegations and police allegations, area_id 1549 (district 1) has the largest number of disciplines per capita.
    - Average household income is higher in this neighborhood
    - Police may respond to these civilian allegations more seriously than anywhere else
    - Population is also lower in these areas so the per capita measurement is higher
- Difficult to determine why the police complaint disciplines are also highest in that area

CP3: Interactive Visualization

Initial Proposal

1. Bar Chart Race (Number of officer/civilian allegation reports in each category per year)
2. Zoomable Circle Packing (Number of officers disciplined together = officers within the same complaint report via data_allegations.crid)

Interactive Visualization

- https://observablehq.com/@jacobkwat/civilian_complaints_barchartrace
- https://observablehq.com/@jacobkwat/police_complaints_barchartrace
- https://observablehq.com/@bodhisattamaiti/zoomable_circle_packing_officers_disciplined

Analysis

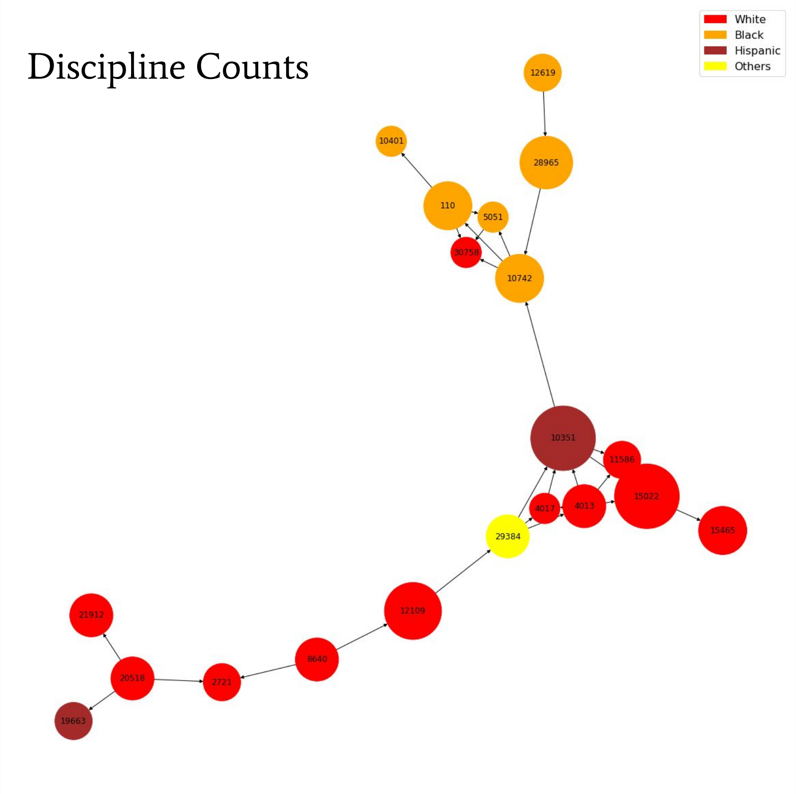
- Bar Chart Races (civilian, police)
    - “Operations/Personnel Violations” is the top category for most years, and usually by a wide margin
    - “Use of Force” has also remained a problem in police conduct throughout the past two decades
    - It is not clear if police reform tactics have been very effective
        - Could be due to lack of data
    - Regardless, focusing on the more prevalent issues in the bar chart may guide reform in the future
- Circle Packing
    - Groups of officers under the same crid can lead to ‘partnerships’ between them in the future
    - This is illustrated more by the graph analysis in CP4

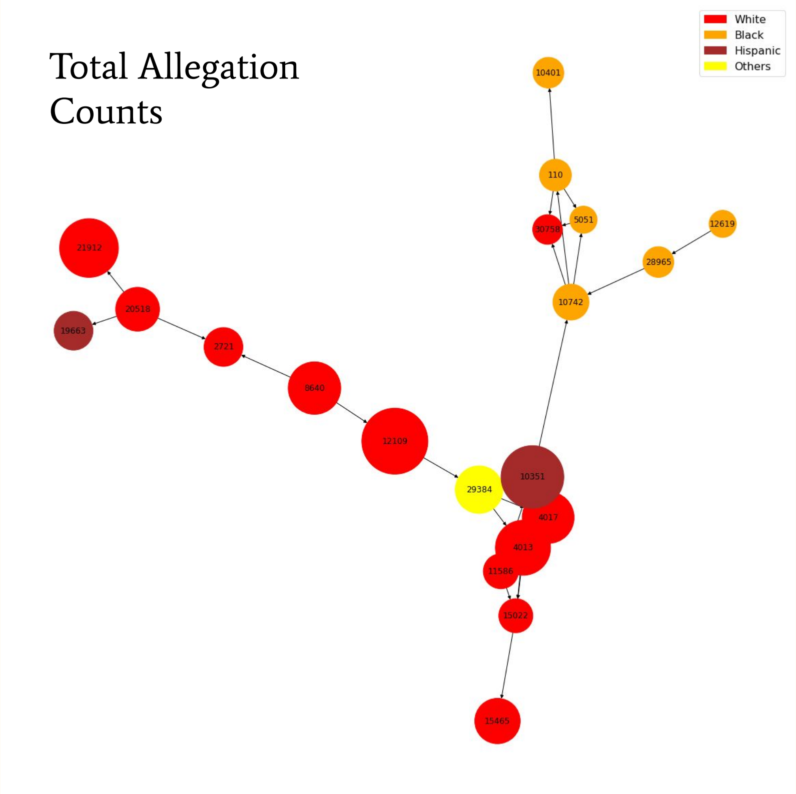
CP4: Graph Analytics

Initial Query

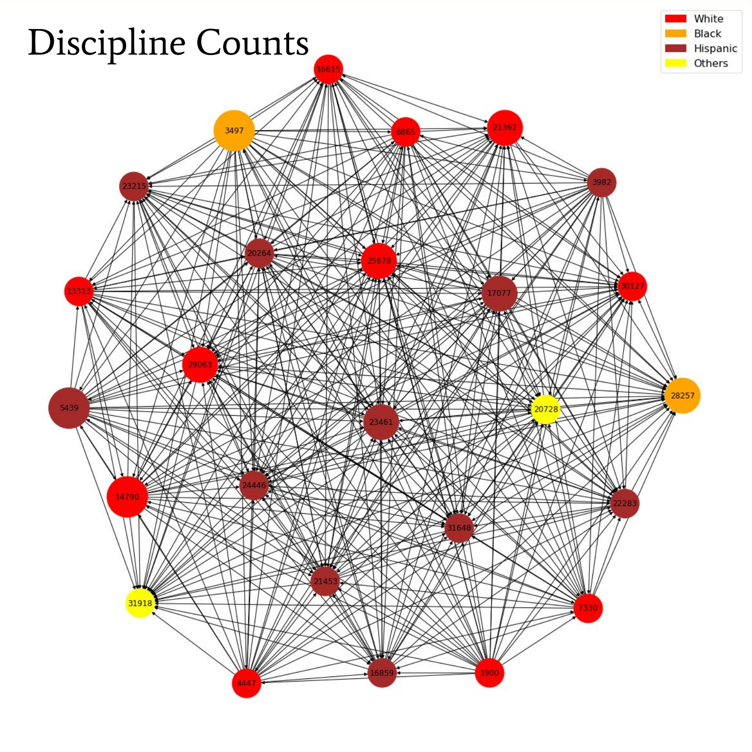
For disciplined officers, we will create a graph where the nodes are the officers and the edges are shared between officers listed on the same complaint. We will create separate graphs for civilian and police complaints, as we've been doing. Finally, we will search for cliques/triangles within this graph to see if there are patterns with which officers are getting disciplined together.

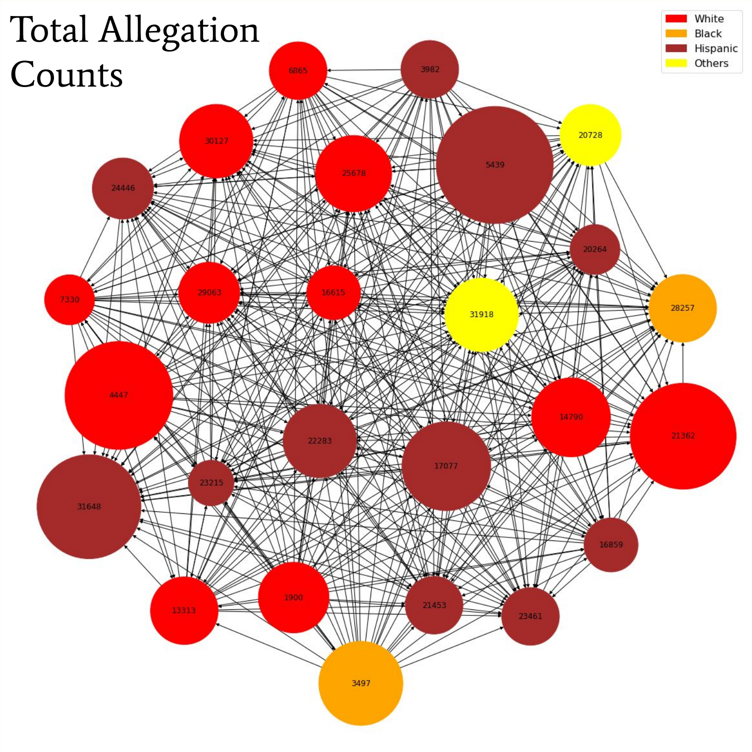
Graph Visualizations - Civilian Component





Graph Visualizations - Police Component





Analysis

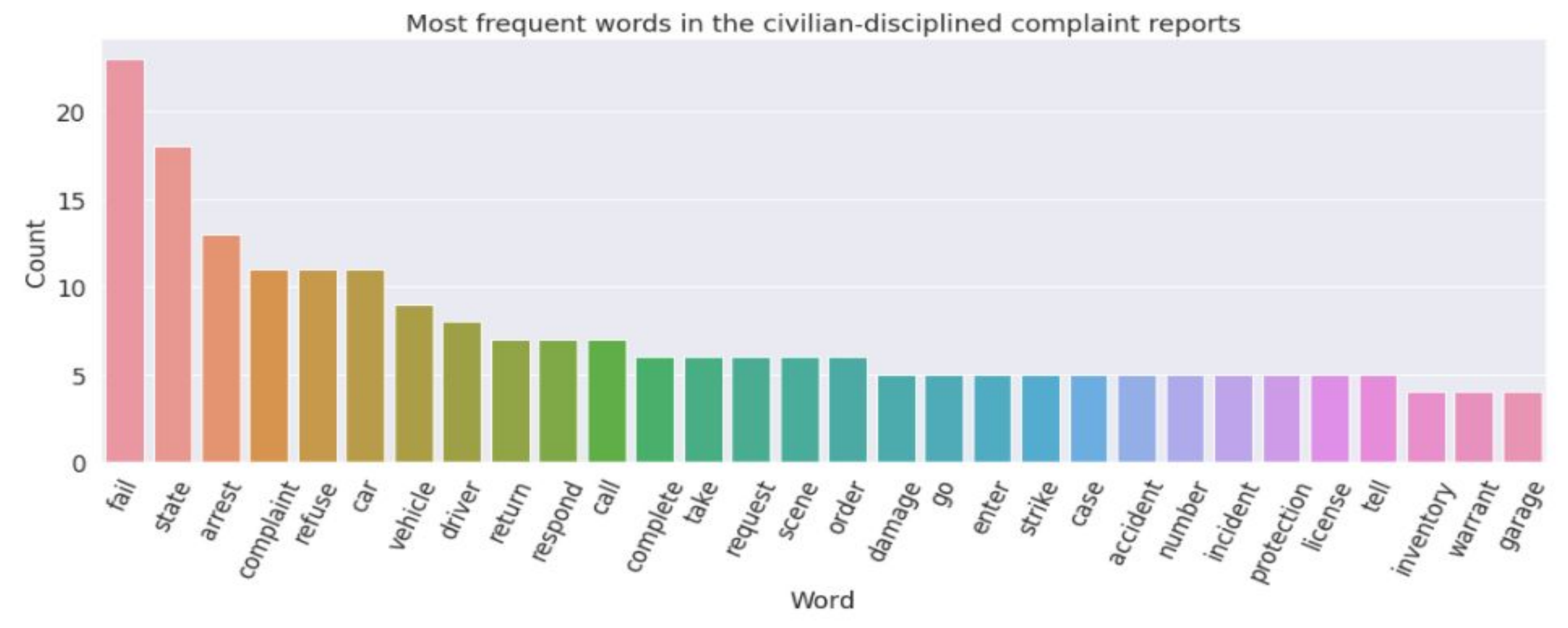
- Civilian Component
  - There is one main group of officers who are white
    - Assumed because the officers who are white have higher allegation counts according to the figure on the right
  - Leads to another group of hispanic officers with a small number of allegations against them.
  - However, there are less allegations amongst the hispanic officers, but they are about equal in discipline counts with the white officers
  - Indicative of a racial bias when disciplining officers or a difference in the severity of complaints when the hispanic officers are involved
- Police Component
  - These officers seem to be extremely interconnected, but fall under a broad range of races
  - We assume that there exists way more connections between officers for officer complaints because of the type of complaints that are usually filed by the department (Personnel Violations, according to checkpoint 3)

CP5: NLP

Initial Query

We're going to use a topic model on complaint report summaries and see if certain civilian report topics lead to officer discipline more than others. Also we will use a keyword based model and see if there is a difference when compared to the topic model. We would like to use this to determine whether a given complaint report summary topic matches with that the reported allegation category is.

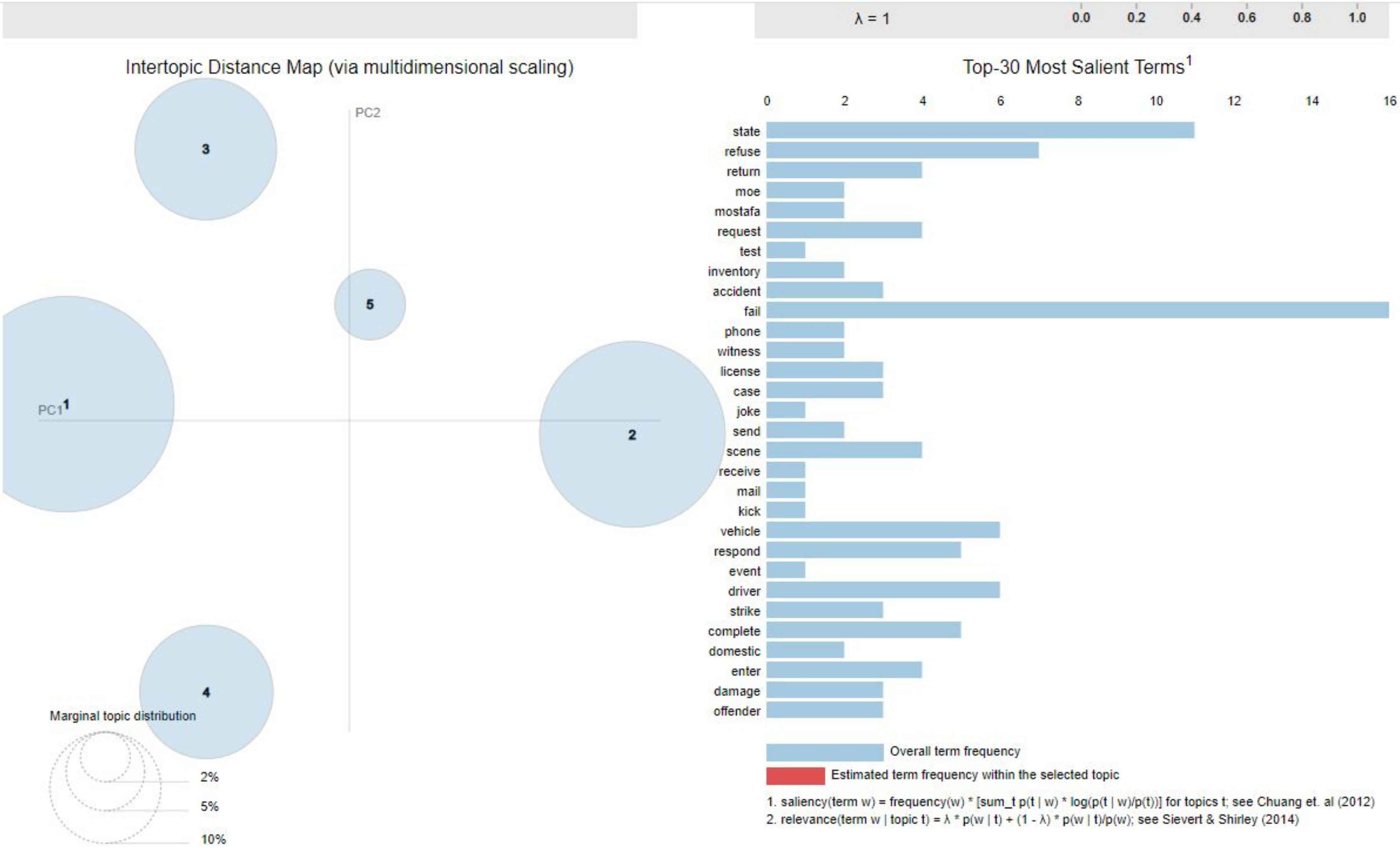
Analysis after processing/cleaning the complaint report texts



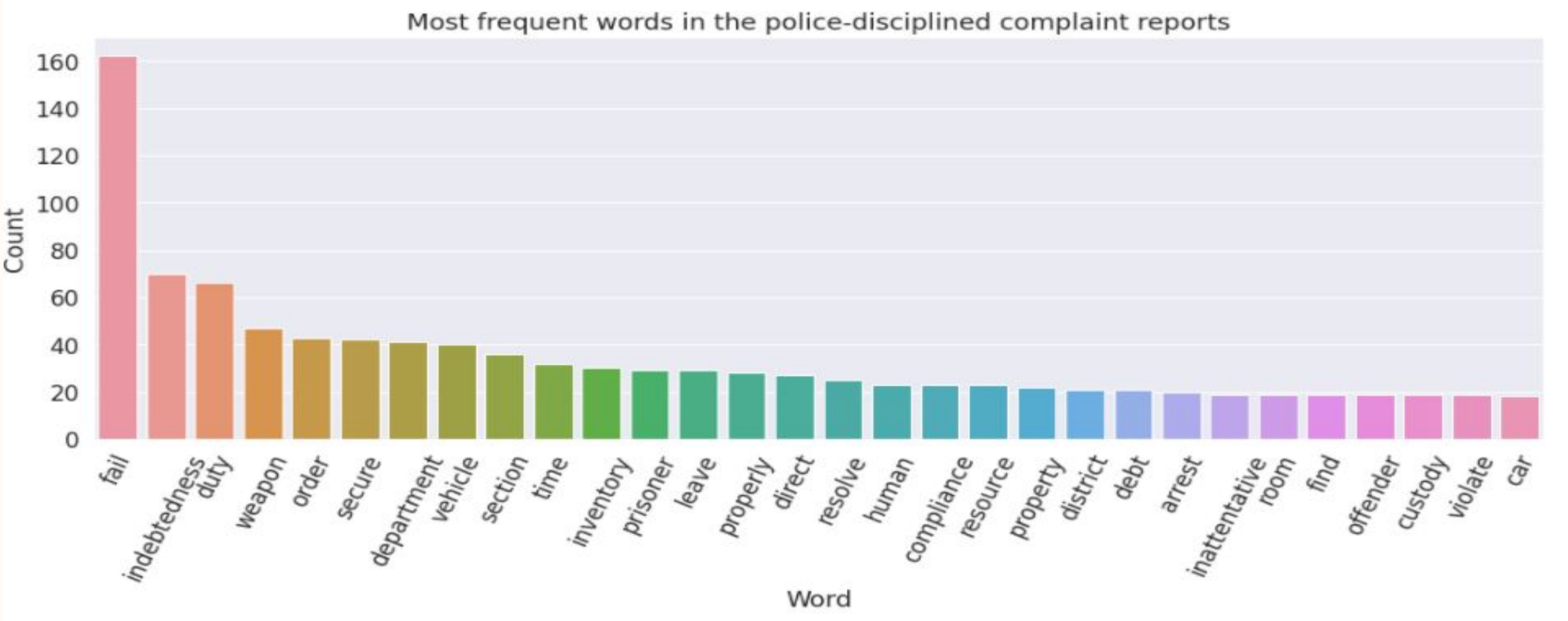
LDA Model Topics for Civilian Complaints (Officer Disciplined)

```
best_ldaModel_civ.show_topics()
```

```
[(0,
  '0.023*"fail" + 0.016*"case" + 0.014*"return" + 0.013*"arrest" + 0.012*"driver" + 0.012*"strike" + 0.011*"domestic" + 0.01
1*"event" + 0.011*"incident" + 0.011*"enter"'),
 (1,
  '0.032*"state" + 0.021*"complaint" + 0.017*"fail" + 0.011*"car" + 0.011*"arrest" + 0.011*"license" + 0.011*"driver" + 0.01
0*"enter" + 0.009*"refuse" + 0.008*"unprofessional"'),
 (2,
  '0.034*"refuse" + 0.020*"moe" + 0.020*"mostafa" + 0.020*"request" + 0.017*"accident" + 0.016*"car" + 0.016*"state" + 0.015
*"phone" + 0.015*"scene" + 0.015*"witness"'),
 (3,
  '0.026*"state" + 0.017*"test" + 0.017*"joke" + 0.014*"return" + 0.014*"inventory" + 0.013*"fail" + 0.010*"receive" + 0.010
*"send" + 0.010*"mail" + 0.009*"kick"'),
 (4,
  '0.033*"fail" + 0.016*"complete" + 0.014*"arrest" + 0.013*"order" + 0.013*"protection" + 0.013*"vehicle" + 0.012*"complain
t" + 0.011*"howard" + 0.011*"franklin" + 0.010*"tell"')]
```

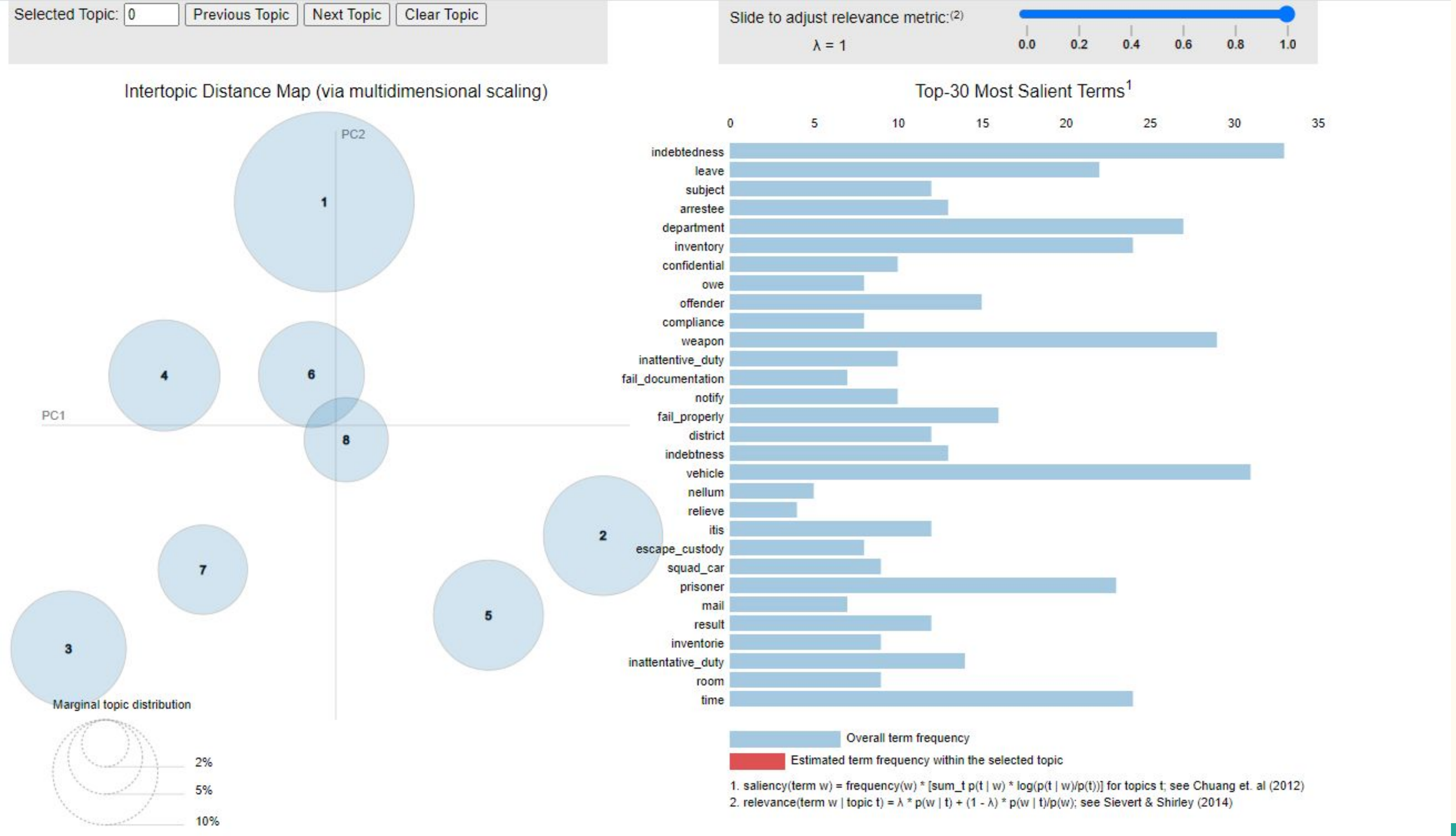

λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
Marginal topic distribution
2%
5%
10%
Top-30 Most Salient Terms1
0
2
4
6
8
10
12
14
16
state
refuse
return
moe
mostafa
request
test
inventory
accident
fail
phone
witness
license
case
joke
send
scene
receive
mail
kick
vehicle
respond
event
driver
strike
complete
domestic
enter
damage
offender
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Analysis after processing/cleaning the complaint report texts



LDA Model Topics for Officer Complaints (Officer Disciplined)

```
[(0,
  '0.029*"leave" + 0.025*"subject" + 0.022*"fail" + 0.012*"relieve" + 0.011*"district" + 0.011*"room" + 0.010*"discover" +
0.009*"2nd" + 0.009*"escape" + 0.009*"security"'),
 (1,
  '0.022*"arrestee" + 0.018*"inventory" + 0.016*"inattentive_duty" + 0.016*"fail_properly" + 0.016*"fail" + 0.015*"time" +
0.014*"prisoner" + 0.011*"escape_custody" + 0.011*"result" + 0.011*"vehicle"'),
 (2,
  '0.026*"fail" + 0.019*"offender" + 0.013*"order" + 0.010*"weapon" + 0.010*"seat" + 0.010*"human_resource" + 0.009*"complia
nce" + 0.009*"inattentative_duty" + 0.008*"division" + 0.008*"specify"'),
 (3,
  '0.016*"owe" + 0.016*"fail" + 0.014*"department" + 0.012*"indebtedness" + 0.011*"time" + 0.010*"duty" + 0.009*"secure" +
0.009*"handcuff" + 0.008*"1s" + 0.008*"prisoner"'),
 (4,
  '0.020*"vehicle" + 0.018*"fail" + 0.013*"secure" + 0.012*"weapon" + 0.012*"state" + 0.011*"nellum" + 0.011*"duty" + 0.010
*"firearm" + 0.009*"number" + 0.008*"fail_properly"'),
 (5,
  '0.043*"fail" + 0.014*"weapon" + 0.012*"arrest" + 0.011*"inventory" + 0.011*"indebtedness" + 0.010*"direct_order" + 0.008
*"medical" + 0.008*"leave" + 0.007*"prisoner" + 0.007*"duty"'),
 (6,
  '0.022*"department" + 0.022*"fail" + 0.015*"indebtness" + 0.015*"confidential" + 0.014*"vehicle" + 0.011*"prisoner" + 0.01
1*"mail" + 0.010*"itis" + 0.010*"secure" + 0.009*"duty"'),
 (7,
  '0.048*"indebtedness" + 0.013*"fail" + 0.013*"notify" + 0.012*"fail_documentation" + 0.012*"compliance" + 0.012*"weapon" +
0.010*"debt" + 0.009*"time" + 0.009*"duty" + 0.007*"park"')]
```


Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
Marginal topic distribution
2%
5%
10%
Top-30 Most Salient Terms1
0
5
10
15
20
25
30
35
indebtedness
leave
subject
arrestee
department
inventory
confidential
owe
offender
compliance
weapon
inattentive_duty
fail_documentation
notify
fail_properly
district
indebtness
vehicle
nellum
relieve
itis
escape_custody
squad_car
prisoner
mail
result
inventorie
inattentative_duty
room
time
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)