Checkpoint 2 Spectacular Cranes

Summary

Our theme is about looking at how the allegation categories have changed over time from previous Chicago Police databases. For now, we are only looking at one old table: `case_info_08282018`. We compare old allegation data (not included in CPDB) with the allegation data found in CPDB. A step we decided to take recently is to see for a given allegation category for a record in the old dataset, what is the most common new category name it has in CPDB? If we can define the "mapping" using this most common idea, then we have some idea about how the category names have changed over time. If we see a record where the mapping is violated (the new category name is not what is usually most common), then we may have some information here. This could reveal some foul play such as changing the allegation category to something less "severe" or something completely unrelated.

For example, if we see that "Illegal Arrest / False Arrest" from the old data maps to "Illegal Arrest / False Arrest" in the new data 98.6% of the time, it is safe to say this is the "true" mapping. In this case, it is fairly obvious because the strings are exactly the same. But this is not always the case. There are a few examples where "Illegal Arrest / False Arrest" has changed to "Use of Profanity," a seemingly unrelated and much less severe category.

To create this mapping, we joined `case_info_08282018` with CPDB tables such as `data_allegation`, `data_allegationcategory`, `data_officer`, etc and used Trifacta to find the mapping and store in two tables: `mapping_allegation_old_to_new.csv` and `mapping_location_old_to_new`.

Questions

1. Is there a significant difference in the distribution of allegation categories in the new data (CPDB) compared to the old data? Visualization of allegation data distribution by category and compare with both versions of allegation tables (density charts)

2. Create a mapping: For each old category name, what is the most common new category name? This will help us establish a mapping from old category names to new category names. (E.g. "Neglect of Duty" records in the old data changed to "Neglect of Duty" in the new data 97.8% of the time. That means it's safe to say that (Neglect of Duty => Neglect of Duty) is the mapping. If we see a record with "Neglect of Duty" in the old data but something different than "Neglect of Duty" in the new data, that might imply the category was changed to something significantly different, which may imply some foul play). We determined the "true" mapping based on which mapping was most common. This can act as a lookup table for the Invisible Inst. or ourselves when we need to know which allegation category a record should correspond to. The goal of this visualization is to allow the user to easily see the mapping. We do the same thing for Location, not just Allegation.

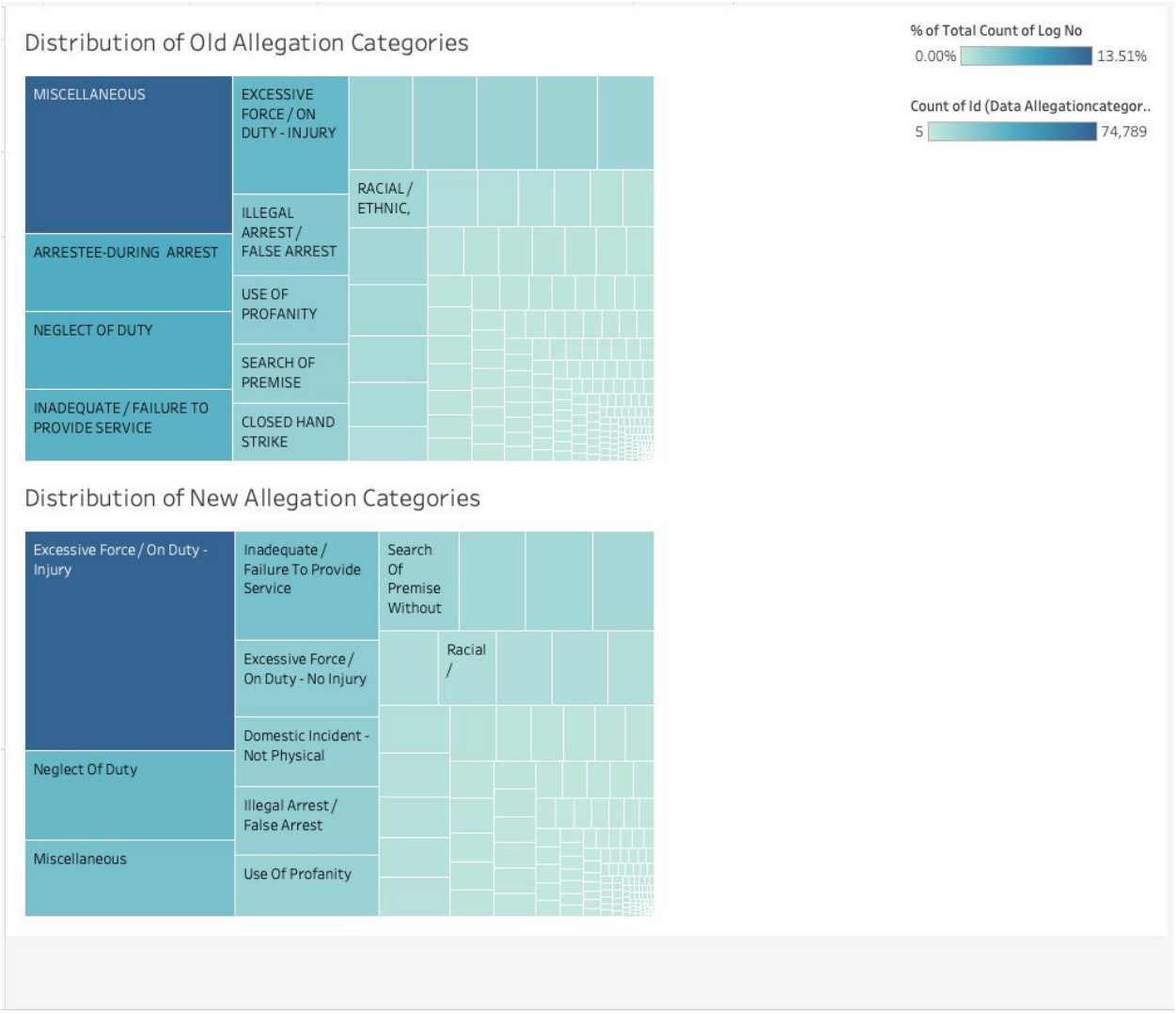
3. Looking at instances where the "true" mapping is violated, try to identify records where allegation category names are commonly changed to a category that is less either "severe" or completely unrelated. This may offer some insight as to whether there are some common techniques used to mitigate the effects of an allegation (i.e. foul play). Like question 2, we also do this for Location and not just Allegation.

4. How have the allegation categories changed from the old data to the new (CPDB)? Are there more or less categories? What is the distribution for the categories, old and new? (pie chart or donut chart)

Findings

1. Go to ‘Question 1’ Dashboard

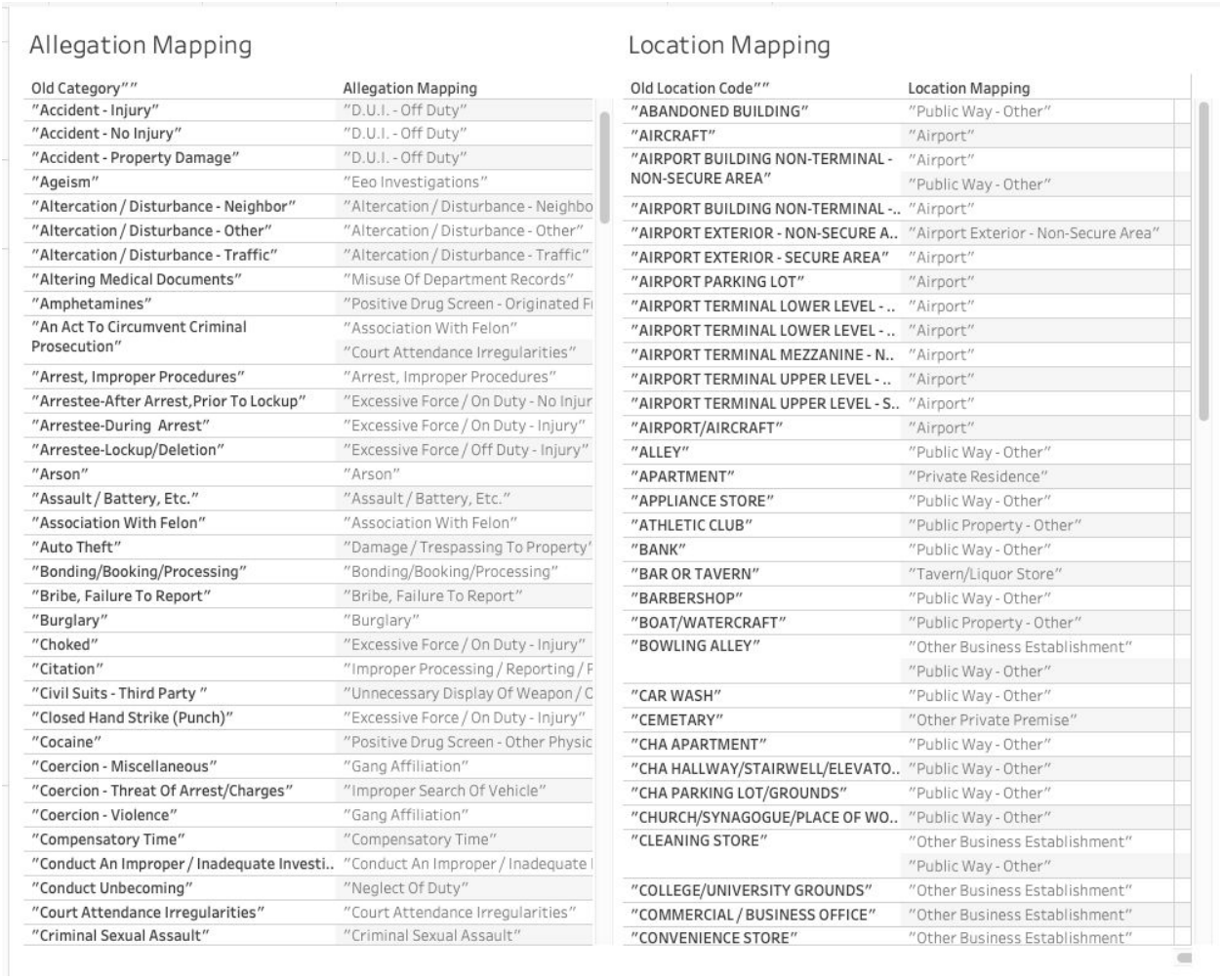
Distribution of Old and New Allegation Categories:



This figure shows the relative distribution of allegations in the old database as well as the new database. In the old database the ‘Miscellaneous’ category is the most frequently used allegation and in the new database it is ‘Excessive Force/On Duty-Injury’. ‘Neglect of Duty’, ‘Use of Profanity’, ‘Illegal/False arrest’ are prominent in both databases. The relative distributions are similar within each database however the ranking of allegations has changed.

2. Go to ‘Question 2’ Dashboard

Allegation and Location Mappings (Most Common):



Allegation Mapping

Old Category""	Allegation Mapping
"Accident - Injury"	"D.U.I. - Off Duty"
"Accident - No Injury"	"D.U.I. - Off Duty"
"Accident - Property Damage"	"D.U.I. - Off Duty"
"Ageism"	"Eeo Investigations"
"Altercation / Disturbance - Neighbor"	"Altercation / Disturbance - Neighbo
"Altercation / Disturbance - Other"	"Altercation / Disturbance - Other"
"Altercation / Disturbance - Traffic"	"Altercation / Disturbance - Traffic"
"Altering Medical Documents"	"Misuse Of Department Records"
"Amphetamines"	"Positive Drug Screen - Originated Fı
"An Act To Circumvent Criminal Prosecution"	"Association With Felon"
	"Court Attendance Irregularities"
"Arrest, Improper Procedures"	"Arrest, Improper Procedures"
"Arrestee-After Arrest,Prior To Lockup"	"Excessive Force / On Duty - No Injur
"Arrestee-During Arrest"	"Excessive Force / On Duty - Injury"
"Arrestee-Lockup/Deletion"	"Excessive Force / Off Duty - Injury"
"Arson"	"Arson"
"Assault / Battery, Etc."	"Assault / Battery, Etc."
"Association With Felon"	"Association With Felon"
"Auto Theft"	"Damage / Trespassing To Property"
"Bonding/Booking/Processing"	"Bonding/Booking/Processing"
"Bribe, Failure To Report"	"Bribe, Failure To Report"
"Burglary"	"Burglary"
"Choked"	"Excessive Force / On Duty - Injury"
"Citation"	"Improper Processing / Reporting / F
"Civil Suits - Third Party "	"Unnecessary Display Of Weapon / C
"Closed Hand Strike (Punch)"	"Excessive Force / On Duty - Injury"
"Cocaine"	"Positive Drug Screen - Other Physic
"Coercion - Miscellaneous"	"Gang Affiliation"
"Coercion - Threat Of Arrest/Charges"	"Improper Search Of Vehicle"
"Coercion - Violence"	"Gang Affiliation"
"Compensatory Time"	"Compensatory Time"
"Conduct An Improper / Inadequate Investi..	"Conduct An Improper / Inadequate I
"Conduct Unbecoming"	"Neglect Of Duty"
"Court Attendance Irregularities"	"Court Attendance Irregularities"
"Criminal Sexual Assault"	"Criminal Sexual Assault"

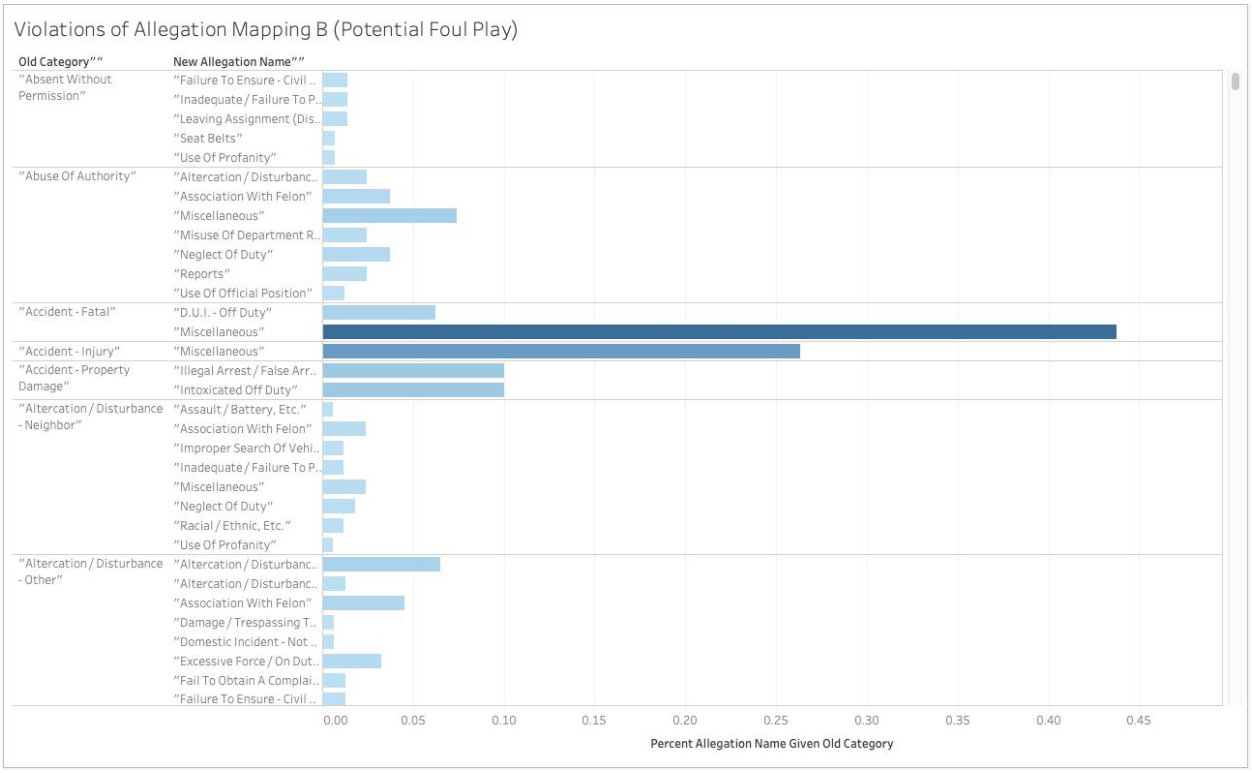
Location Mapping

Old Location Code""	Location Mapping
"ABANDONED BUILDING"	"Public Way - Other"
"AIRCRAFT"	"Airport"
"AIRPORT BUILDING NON-TERMINAL - NON-SECURE AREA"	"Airport"
	"Public Way - Other"
"AIRPORT BUILDING NON-TERMINAL -..	"Airport"
"AIRPORT EXTERIOR - NON-SECURE A..	"Airport Exterior - Non-Secure Area"
"AIRPORT EXTERIOR - SECURE AREA"	"Airport"
"AIRPORT PARKING LOT"	"Airport"
"AIRPORT TERMINAL LOWER LEVEL - ..	"Airport"
"AIRPORT TERMINAL LOWER LEVEL - ..	"Airport"
"AIRPORT TERMINAL MEZZANINE - N..	"Airport"
"AIRPORT TERMINAL UPPER LEVEL - ..	"Airport"
"AIRPORT TERMINAL UPPER LEVEL - S..	"Airport"
"AIRPORT/AIRCRAFT"	"Airport"
"ALLEY"	"Public Way - Other"
"APARTMENT"	"Private Residence"
"APPLIANCE STORE"	"Public Way - Other"
"ATHLETIC CLUB"	"Public Property - Other"
"BANK"	"Public Way - Other"
"BAR OR TAVERN"	"Tavern/Liquor Store"
"BARBERSHOP"	"Public Way - Other"
"BOAT/WATERCRAFT"	"Public Property - Other"
"BOWLING ALLEY"	"Other Business Establishment"
	"Public Way - Other"
"CAR WASH"	"Public Way - Other"
"CEMETARY"	"Other Private Premise"
"CHA APARTMENT"	"Public Way - Other"
"CHA HALLWAY/STAIRWELL/ELEVATO..	"Public Way - Other"
"CHA PARKING LOT/GROUNDS"	"Public Way - Other"
"CHURCH/SYNAGOGUE/PLACE OF WO..	"Public Way - Other"
"CLEANING STORE"	"Other Business Establishment"
	"Public Way - Other"
"COLLEGE/UNIVERSITY GROUNDS"	"Other Business Establishment"
"COMMERCIAL / BUSINESS OFFICE"	"Other Business Establishment"
"CONVENIENCE STORE"	"Other Business Establishment"

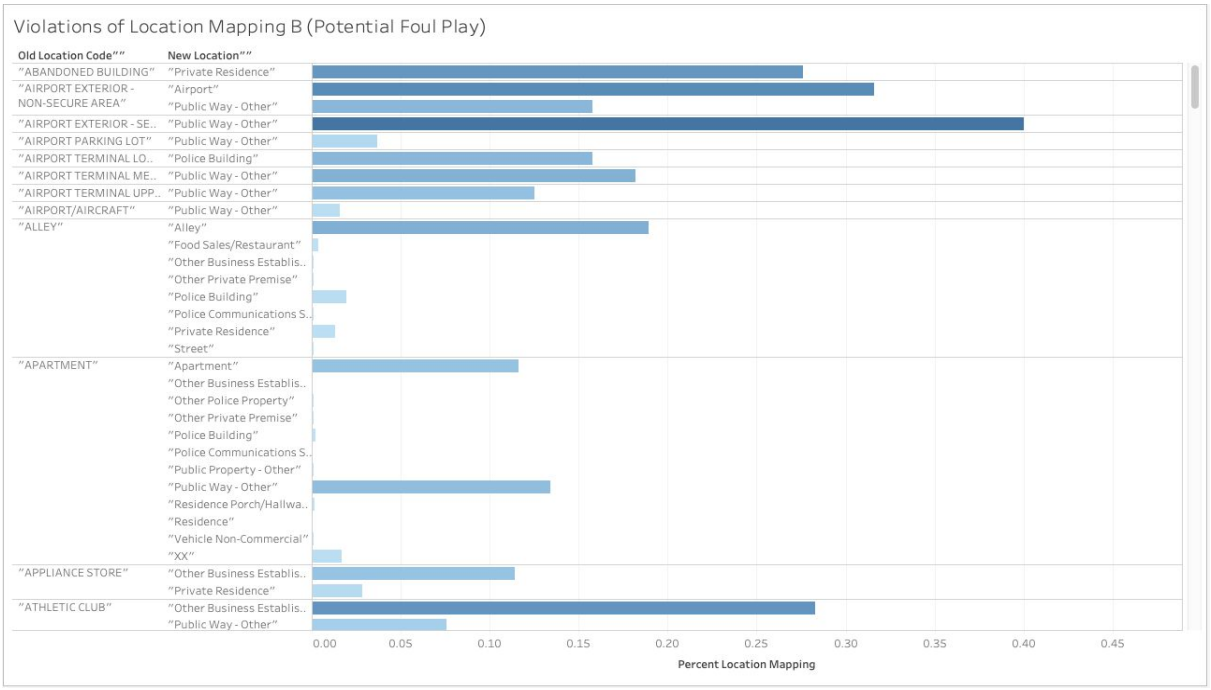
This figure shows a mapping of each allegation in the old database to the most frequently used allegation in the new database. In other words, ‘Choked’ in the old database was most frequently mapped to ‘Excessive Force/On Duty - Injury’ in the new database. Similarly for location, ‘Bar or Tavern’ was most frequently mapped to ‘Tavern/Liquor Store’. This mapping serves like a ‘look-up’ table of sorts to make sense of what allegation/local labels in the new database most likely represent in the old database. In light of the fact that an old allegation can map to several new labels in the new database, this table provides a reference for what might be defined as effectively ‘unchanged’ in the new database. Many of the allegations map to the same allegation in the new database but there are a significant number of changes in terminology. Some don’t make obvious sense ‘Not at home’ to ‘Seat Belts’’. In fact, all of the following map to ‘Seat Belts’ : Not at home, Medical Roll Violation, Failure to Provide Documentation, Failure to Appear at Mss, Failure to receive approval for travel, timely notification to unit, and working while on the medical /IOD.

3. Go to ‘Question 3AllegationA’ and ‘Questions3AllegationB’ and ‘Question3LocationA’ and ‘Question3LocationB’

Violations of Allegation Mapping:



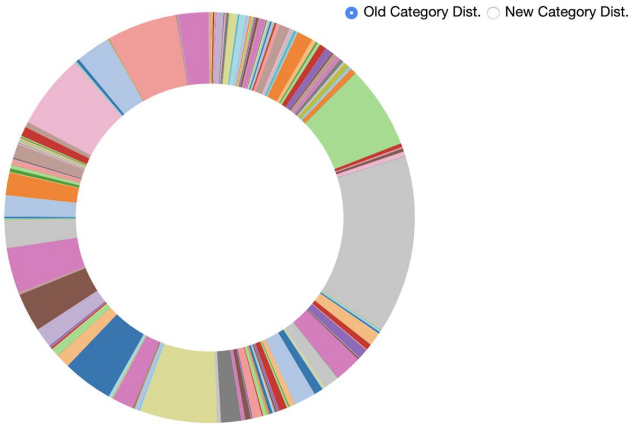
Violations of Location Mapping:

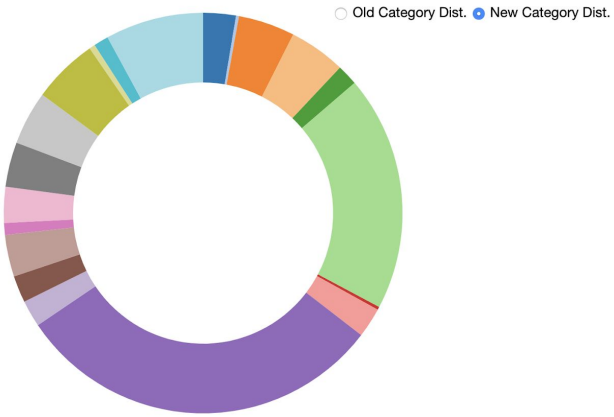


These figures show us visual representations of the “other” mappings for each allegation and location. In other words, while in Question 2, we focus on the most common mapping and use that to define what it means to be “unchanged”, in this part we are focusing on all the “changed” mappings. For each (allegation and location) there is an ‘A’ and ‘B’ representation. They both display the same distribution of mappings for each element in the old database but do it in different ways (horizontal vs. vertical). It’s interesting to see how certain allegations/locations seem to map to a myriad of things in the new database (for instance ‘Altercatio/Disturbance Neighbor’).

Understanding the findings in Part 2 and Part 3 gives us a stepping stone for secondary analysis and looking at what may be underlying these patterns of changes (for instance, the officers involved).

4. D3 Visualization, which was not updated for re-submission.





Note: Tabs titled “(IGNORE)” are from our first submission. Please ignore.