POST project documentation  
Publication: Oct 1-4, 2017  
Jennifer Bjorhus, MaryJo Webster, Dave Hage(editor)

Series of stories published in 2017 on police officers convicted of misdemeanor or higher-level offenses while they held a valid police license. Stories focused on how the state’s licensing authority has very narrow window to revoke an officer’s license or require some other form of discipline. The system is set up to rely on police chiefs and sheriffs to deal with problematic officers.

Data sources:

1. From the POST Board: Data on all disciplinary actions since 1995, including name, license number, and details about the disciplinary action such as type of action and date it took place. Unfortunately, the reason for the action was usually something vague like “felony level offense”
2. From the POST Board: Data on all licensed peace officers since the state’s licensing system began in 1978, including full name, date of birth, license start date, expiration date and current status. Later we asked for a supplement that listed dates that licenses had been switched to “inactive” for those that had a current status of inactive. Ultimately, we whittled this down to only officers who had a valid license as of 1/1/1995 or later.
3. We also used court conviction data, which we obtain each month from the state court system (MNCIS) at no cost.
4. We requested 2016 arrest data from the St. Paul and Minneapolis police departments (only St. Paul fulfilled our request within the time frame we requested)
5. We also had jail bookings for all jails in the state going back to the 1990s. We obtain that on a yearly basis from the Minnesota Department of Corrections at a cost of about $30.
6. We also obtained data, at no cost, from Matthew Hickman at Seattle University on the number of revocations in each state in 2015. We paired that with data from the FBI (also no cost) on the number of licensed peace officers in each of those states, in order to generate license revocation rates for each state.

The license data from the POST board included some cases of multiple records for the same person. These turned out to be primarily people who initially had a part-time license and then switched to a full-time license.

In order to join to crime records properly, I needed a list of licensed officers that only had one record per person. So I flagged all the part-time records and also flagged as “not using” any that were people who had another full-time records (this leaves some people in here who only have a part-time record).

I also marked as “not using” records where the license expired prior to 2000 – since our crime searches aren’t going to go back that far.

I also created 2 new lastname fields – lastname\_clean and lastname\_clean\_2

The first one is a cleaned up version of the original; stripped of suffixes, commas, periods and apostrophes

The second is to deal with people who might have their name listed different ways. For example, in the first field it might be “O BRIEN” and in the second it will be “OBRIEN”

There were some that I left that I’m certain will have a space – such as “ST MARTIN” or “VANDER CLEEVE”

For people with double names – either hyphenated or not – I used the second last name field to put in just the first half of their double name.

COURT DATA:

We had two batches of court data. The first is a monthly file we get from MNCIS that has all the convictions that are in the MNCIS system (which started in the mid-2000s). Each month we get access to a new file that replaces all the past data (this is so that expungements are removed). That file only contained convictions, with one record for each charge the person was convicted on. The second is an archived file that we kept from before they switched to MNCIS. This had court records going back into 1990s, and it contained all charges (convictions, dismissals, etc), including cases where the entire thing was dismissed. The court refused to provide us the dismissal data for the newer cases, citing their rules for access to bulk data.

I spent considerable time breaking apart the court data to weed out traffic tickets and DNR violations that we didn’t care about. We chose to include only convictions for misdemeanor or greater, from 1995 to present.

We found a handful of cases where the conviction was a misdemeanor but the judge chose to sentence the person at the petty misdemeanor level. We decided to not include those in our count of convictions.

I didn’t spend much time cleaning up names in here, other than getting rid of leading/trailing spaces and turning it all to uppercase, getting rid of periods and putting the suffix in a separate field.

MATCHING DATA:

The crux of our work was to do an enterprise join between the police licensing data and the court conviction data. The licensing data included about 20,000 people who had held a license at some point since 1995 (or currently held one). The court data, combined with jail bookings data, was more than 4 million records – so I eliminated case types we didn’t care about (traffic tickets, DNR violations, etc). Then I assigned each person in the remaining court data a unique ID and then did a group by query on the unique ID, name fields and date of birth --- this boiled it down to about 1.4 million records.

We put those two sets of names into dedupe.io and told the program to find the matches. The program actually finds matches across the two datasets and also within a dataset. So it spent a lot of time finding matches just within the court/jail table. This yielded nearly 1,000 matches. There were a few bad matches where it clearly wasn’t the same person (different middle initials or names, for example) that we threw out right away.

But then we needed to do the second step – which is compare their conviction date with the date range they had a valid license. The number of people got cut in half since many of the convictions were either before or after their time as a police officer.

We did that matching in mySQL. We initially based this on the date of the case was filed (which was consistently provided in the data). But later Jennifer found out that the POST board only takes action if the license is valid when the officer is SENTENCED. So we got sentencing dates for anybody that we didn’t have one for and then checked to see whether their conviction occurred while the license was still valid. We chose to keep the ones that had expired before sentencing (but case was filed while they were active) and we included them in the final tally because the executive director of the POST board expressly said that one reason that they didn’t discipline all officers with a conviction was because of these licenses expiring. (Typically if the license expires and is not renewed, it’s an indicator that the person lost their job in Minnesota)

After pulling the appropriate convictions for all the officers, we put that data in Microsoft Access and started building out a custom table there. Jennifer had things she wanted to add (like a flag indicating if the case involved domestic violence) and we needed to classify cases and make notes.

Through this process we learned that the bulk data we get from MNCIS doesn’t include all the fields that you see on a MNCIS report that you pull up on their website. Key fields that are missing are the “level of sentencing” (this indicates if the judge sentenced the person at a level lower than what they were convicted of – i.e. sentence as petty misdemeanor instead of misdemeanor). This ended up being important because the POST board bases their decision on that level of sentencing.

Eventually we sat down and spent about three days going through each of the more than 600 cases (some officers had more than 1 case) and we checked our data on key points – including that level of sentencing, the sentencing date, higher level charges that were dismissed, etc. We compared our data to the MNCIS reports that we printed out from the court’s website. This turned out to be brutal work, but definitely worth our time.

Once that was complete, we started the analysis. Since our table included each “charge” that resulted in a conviction for each case, I created a field that flagged the highest-level charge in each case. Then I used that as the primary record for subsequent analysis. All the counts on things like number of DWI’s or number of felonies are based on that case-level analysis.

I put a copy of the final data in PANDA for storage -- it’s called “police officer convictions 1995-2016”. There are also the “police licenses” and “POST board discipline 1995-2016” in PANDA. I’ll put copies of all three with this documentation on github, as well.