

do list:

Clean up R code and put on github

Story about our journey requesting all these records?

JOURNAL:**Dec 14:**

The last of the Anoka files arrived today.

Nov 14:

Last batch of Minneapolis data arrived! If we enter the remaining cases we have now (about 65), we'll surpass 1,400. Still waiting on files from Anoka, but might not be worth entering all of those. We fell short of the 2,000 we expected due to a variety of factors. Among them -- St. Paul holding "open" a lot more cases than we expected; and the out state agencies giving us the child cases we didn't want; and Minneapolis also came in lower than we estimated, but that was always an estimate that we knew would be hard to pin down because we don't know how many cases they have left open. Based on the fact that they report about 400 cases to the BCA each year, and we got about 650 cases from 2015 and 2016, that means there are something in the neighborhood of 150 cases sitting open from those 2 years.

The 2017 sentencing data arrived. Anne Wall sent the SPSS file, like she did the last time. A few of the variable names were different (Inctype instead of inctype), so I had to deal with that in my code.

Nov 13:

Anne Wall sent me a new data file of sentencing data, this time including the field that indicates how many prior criminal sexual conduct convictions each person had. It works out to just 5% of all the cases (the 13 and older victim cases)

Talked to John Elder about why we don't have the last batch from Minneapolis yet. Kyle MacDonald had told me 3 weeks ago that it was finished and it just needed to be reviewed by the PIO office. John said he paid someone overtime this weekend to get it done. He said it got delayed cause his office is working with just 2 staff people, when there should be 4, plus they've had a lot of big news events (officer-involved shooting, etc). He said they found a couple questionable things they are chasing down today, but expect to send it off tomorrow.

Nov 12:

If our goal is to explain why so few CSC offenders get prison time, then we have to explain the confluence of factors playing into that:

1. The majority of cases are statutory (60%), which rarely carry a prison sentence (unless the person has a significant criminal history, which doesn't happen very often). Just 26% of statutory cases had a presumed prison sentence.
2. The majority of offenders (60%) have a 0 criminal history score. (Even if you split it out, 55% of non-statutory cases have 0 criminal history; 64% of statutory cases). 82% of offenders have a score between 0 and 2.
3. Excluding the statutory cases, we also see judges are more likely to give a downward departure (jail instead of prison) to an acquaintance case than other types of cases (stranger, family, person of authority), even when the offenses fall at the same spot on the sentencing grid. (same severity, same criminal history score). The majority of these downward departures are supported by prosecutors, according to the data, and we should be able to anecdotally prove that almost all come about because of plea deals.

Important to note:

The sentencing grid treats repeat sex offenders much more harshly than even repeat felony offenders on the main grid. They move up the criminal history score ladder faster with a repeat sex offense than if their prior offense was any other type of felony. Even if an offender is convicted of an offense that doesn't carry a presumed prison sentence, a prior sex offense will automatically push them into a presumed prison sentence. However, as we noted in story #2, cops aren't doing that great of a job catching the repeat offenders.

When sex offenders are released from prison after serving 2/3 of their sentence behind bars, they have to serve the remaining 1/3 on probation, but then they also get another probation term that goes beyond that.

Nov9:

Felony-level criminal sexual conduct convictions, charged under any of the state laws and involving any age victim (about 2,198 cases):

28% of acquaintance cases resulted in a state prison sentence

64% of stranger cases resulted in a state prison sentence

Felony-level criminal sexual conduct convictions, charged under statutory provisions, victims ages 13-17 (1,321 cases):

18% of acquaintance cases resulted in a state prison sentence (almost all of these are assailants with a significant criminal history score)

23% of stranger cases (there are only 43 of these cases)

Felony-level criminal sexual conduct convictions, charged under laws that require proof of force or incapacitation (875 cases):

*these tend to be victims 18 and older, but it is possible for a prosecutor to charge a case under these statutes with a victim of any age

47% of acquaintance cases resulted in a state prison sentence

74% of stranger cases resulted in a state prison sentence

Among these cases, judges gave either a downward dispositional departure (meaning jail, not prison) or a downward durational departure in 30% of the acquaintance cases and only 22% of the stranger cases. The majority of the departures for acquaintance cases were dispositional, while the departures on stranger cases were about half durational and half dispositional.

More specifically, the head of the sentencing guidelines commission recommends that we focus on cases that have the same severity level and criminal history score on the sentencing grid. The most commonly used spot on the grid is criminal sexual conduct 3 (which is a severity level of C), with a criminal history score of 0 or 1. These cases would all have a PRESUMED prison sentence, of a minimum of 41 months (keep in mind that time served behind bars would only be two-thirds of that).

If we look at the force/incapacitation cases (excluding statutory) that fall at this level on the grid, 47% of acquaintance cases got a prison sentence and 61% of stranger cases did. This shows that, even in cases with very similar qualities as far as the sentencing grid is concerned, the acquaintance cases are still more likely to get a departure. A data analyst at the sentencing guidelines commission, who has helped me with this analysis, tried to explain why this is happening by essentially saying these cases are “complicated” and the circumstances really vary widely from case to case.

Nov 8:

Things we wish we had tracked in the database:

- 1) On the victim cooperation question -- separate the ones where she told the cops she didn't want to pursue charges or didn't want to continue, versus those where the cops just couldn't reach her/find her
- 2) Something about the age of the victim. Since they are redacted, this is hard. But there are clues that tell you whether the victim is still in high school, or possibly in college. Reports on teens tend to include mentions of parents or a high school. Those that mention the victim was at a bar or a club can be assumed to be at least over 18. Would be helpful to at least be able to separate teens versus those 18 and over.
- 3) The MNCIS case number and the specific statute(s) charged under for cases that were charged and/or convicted. We realized late in the game that the bulk of cases resulting in conviction were “statutory” cases, involving teen victims and being charged under lenient statutes that allow defendants to use consent as a defense.
- 4) Did the suspect have a prior conviction or is there any mention of the suspect having been implicated in other sexual assaults?

- 5) Regarding acquaintance versus stranger would also like another level of detail -- that they had been romantically involved either at time of incident or at some point in the past; that they had known each other for at least 2 weeks; that they knew each other for less than 2 weeks; or that they had just met that day. We also wondered how to deal with assaults made by Uber drivers? We ended up considering them stranger cases because it was a professional relationship.
- 6) We put down the name of the suspect(s) and used that to determine whether police had a named suspect. However, sometimes these names showed up late in the investigation. I'd rather have a field that indicates yes/no -- did the victim name the suspect at the time of the initial report?

Nov 1:

Notes from analysis of sentencing data. Nate Reitz, head of the sentencing guidelines commission, recommended that we exclude statutory cases and only compare cases sentenced at the same statute and same severity (ie. CSC 3 -Force cases with severity=C).

In the past decade, less than one-third of people convicted of felony-level criminal sexual conduct, involving victims age 13 and older, were sentenced to a year or more in state prison. The remainder were sentenced to less than a year in county jail combined with probation, or in some cases, only probation.

This is partly driven by the fact that a conviction for the most common kind of sexual assault – what's often referred to as an acquaintance rape – is much less likely to result in a prison term than those involving a stranger assailant, even when comparing cases charged under the same state statute, according to a Star Tribune analysis of about a decade worth of felony sentences compiled by the Minnesota Sentencing Guidelines Commission.

For example, most rape cases involving penetration and either the use of force or a situation where the assailant knew the victim was too incapacitated to consent are charged as third-degree criminal sexual assault with a severity level that calls for a sentence of at least 41 months in state prison. Sentencing data shows that these cases result in a jail sentence of less than a year, or even probation, 43 percent of the time when the assailant was an acquaintance of the victim, compared to 26 percent of the time when it was a stranger.

This excludes cases with victims under age 13 and those referred to as statutory rapes, where the victim was between ages 13 and 18 and the assailant was older. In these cases, the prosecutor can charge the assailant without having to prove the victim didn't consent. Only about 12 percent of those cases result in a prison sentence.

Side notes:

The pattern holds for CSC 1-force/incapacitation cases, with a severity level of A, which the sentencing grid calls for a minimum of 144 months in prison. (there aren't enough CSC 1 cases with a B or C severity to use those)

Acquaintance cases got sentenced to prison 86% of the time; reduced to jail or probation 14% of the time

Stranger cases: 96% of the time to state prison; reduced to jail or probation 4% of the time.

As I mentioned in my earlier email, an insane share of CSC 3 cases are statutory ones. If you broaden that out to the entire universe of felony CSC involving someone age 13 and older, statutory cases account for 66 percent of all cases!! These rarely get a prison term. That goes a long, long way in explaining the 70-30 split. Perhaps we need to mention that?

Oct 25:

Our lawyer, Randy Lebedoff, sent a letter to the Itasca County Sheriff about the fact that nine months after making our request, we still don't have any of the 74 reports we requested

I sent another check to Kelli Hill for the third batch of Anoka records. Expecting two more batches.

Oct 22:

Kyle at Minneapolis PD says he has finished the last batch of records and is waiting for the PIO (and perhaps others?) to review that batch, plus to make sure he didn't miss anything. (update: as of Oct 30, still waiting for this last batch)

Payments for data so far:

Anoka, July 13: \$276.47

Anoka, Sept 18: \$238

Anoka, Oct 22: \$467.50

Cass County, \$38.56

Duluth: \$250+103

Mankato: \$101.25

Moorhead: \$232.50

Rochester: \$100

Shakopee: \$29.95

St.Cloud: \$221.33

Wright County: \$168

St. Paul: \$951

Minneapolis: No record of them charging us anything

No cost: Brainerd, Bloomington, Brooklyn Center, Brooklyn Park, Grand Rapids, St. Louis County

Oct 10:

Rochester Police agreed to go back through the cases they sent us and give us a spreadsheet with the suspect names. In half or more of the cases, the PD had redacted suspect names, without a legitimate reason to do so. (juvenile suspects can be redacted). They sent the spreadsheet, with case ID number, suspect name and what they knew about the outcome of the case (sent to prosecutors, charges declined, etc). I put the names into AirTable. In the process of doing that I found some cases in OpenSemantic that we had missed, but for some weird reason they weren't tagged as "Needs review" any longer. Abby and Jennifer went back through and got those entered.

Sept 18:

Kelli Hill from Anoka County says she's been authorized to work overtime in order to work on our request. They are way behind schedule. I requested a check be cut and mailed to them to cover cost for the batch she is finishing up this week. She won't send the files until we send the money, unfortunately.

August 15:

Yesterday in an interview with retired Minneapolis detective Mike Sauro, he was talking about redlining cases and he kept saying, essentially, that a case was more likely to be redlined if there wasn't physical evidence collected. So I ran some numbers. I created a new field that combined whether or not there was a rape exam and whether or not other physical evidence was collected. (analysis is in the "rape2018.rmd" markdown page)

These are for all our 1,100+ cases:

Of the cases with both a rape kit and other physical evidence, 90% were assigned to an investigator

Cases with other evidence, but no rape kit, 86% were assigned

Cases with only a rape kit and no other physical evidence, 69% were assigned

Cases with neither a rape kit, not other physical evidence, 66% were assigned

For Minneapolis only:

Both kit and other evidence: 91%

Kit only: 68%

Other evidence only: 88%

No kit, no evidence: 61%

One caveat to keep in mind is that we noted whether there was physical evidence, regardless of whether it was collected by the patrol officer or the investigator. However, we've been told that this evidence generally needs to be collected by the patrol officer because it might not be available any longer by the time the investigator starts work on the case (if they are assigned at all). Unfortunately in our data, we didn't note who collected the evidence.

August 14:

I put the Choi data in R (markdown file called rcas.rmd) and looked at whether the outcomes were different depending on the victim's race. I collapsed the race field to "white" or "person of color." It showed absolutely zero difference.

They were able to capture the victim race in 60% of the cases, which is good (but not great) And it works out to be about 29% people of color and 31% white (and then 40% unknown) – which seems abnormally skewed to people of color. I'm guessing, then, that the unknown group is primarily white people.

The top-level outcomes show:

12% of cases involving a victim of color were charged; Same percentage for cases involving a white victim; unknown category is 11%

15% of cases involving a victim of color were declined; Same percentage for cases involving a white victim; unknown category is 27%

64% of cases involving a victim of color were not even referred to Choi's office; 67% of cases involving a white victim; unknown category is 51%

I also looked at the share of cases referred to Choi's office overall and it's too close to say anything – 26% of cases involving a person of color, 27% of cases involving a white victim; 35% for that unknown category

The unfounded rate is also too close – 4% of cases with a person of color; 6% with a white victim (really small number of cases in both instances)

I looked at some police investigative work categories, as well.

Victims of color were interviewed 40% of the time; white victims were interviewed 38% of the time. (that unknown category is 36%)

Investigators requested names of witnesses that would corroborate the victim's story 35% of the time with a victim of color; and 39% with a white victim; 39% for that unknown category. This one shows a slight difference, but not enough to make anything of it.

August 13:

Met with Kaarin Long, the assistant county attorney in Ramsey County who did the case review for their office. She walked me through all the data fields and gave me a jump drive with the data, a “key” (record layout), and some other documents that Brandon had requested. The data is very detailed. They redacted a few things -- DOBs of the victim and suspect and house numbers from any addresses. Unfortunately they didn’t include the ages of the victim and suspect. But it has a ton of other useful stuff including race of both victim and suspect, more details about victim intoxication and the relationship between the two (did they know each other for more than 2 weeks? Or just met that day? Or in a relationship?). They also logged whether the victim requested no further action on the case (something we really wish we had done in our data).

Update on cases we’re waiting for:

Bloomington – 5 new cases came in last week. She says there is one more batch (not sure how many but it’s probably about a dozen; I told her it was ok to exclude child cases)

Rochester – They seem to think they sent us everything, but I can only find 34 PDFs from them. I’ve given them 51 case numbers for adult-victim cases that we still need. Hoping we didn’t do something wrong in the transfer of files.

Wright County – 9 new cases sent last week. Waiting for 50 others, however they will exclude child cases from those.

Itasca County – No word. We requested 72 cases; haven’t gotten any.

St. Cloud – Waiting for 25 cases.

Anoka County – We requested about 458 cases (I think); have received 57 cases. I’m going to tell them to exclude child cases and maybe that will speed things up (and reduce the number of cases). Kelli has been out of the office due to family medical situation.

Minneapolis – We’ve logged 541 cases; there are a couple dozen waiting to be read. We’re getting about a dozen new files each week. No idea when it will be finished.

I’m also circling back to Duluth PD, St. Louis County, Moorhead and Grand Rapids to make sure the number of adult cases we have logged is a viable percentage of their total closed adult cases. So many of the files we got from these agencies weren’t viable because they were child or family cases or unfounded.

Aug 9th:

Talked to Anne Wall at the Sentencing Guidelines Commission. She helped clarify how I can split up the sentencing data on CSC cases to remove those cases that were convicted under statutes that only apply to child victim cases. This doesn’t necessarily mean we won’t have any

cases with child victims, but it should winnow the majority of them out. She is out of the office next week, so any further questions will need to wait until after that.

Aug 3:

Other media reacting to our stories:

<http://www.herald-journal.com/archives/2018/stories/sheriff-candidate-forum.html>

http://www.grandrapidsmn.com/opinion/columnists/sexual-assault-cases/article_093efcfe-9506-11e8-a75a-e31494a4f04d.html?utm_medium=social&utm_source=email&utm_campaign=user-share

<https://www.kare11.com/article/news/behind-the-star-tribunes-rape-investigation/89-576830607>

<https://www.kare11.com/video/entertainment/television/programs/breaking-the-news/star-tribunes-rape-investigation-part-ii/89-8201677>

<https://www.mprnews.org/story/2018/07/26/sexual-assault-investigation-prosecution-free-man>

<https://www.tpt.org/almanac/video/denied-justice-special-star-tribune-series-31114/>

<https://www.mprnews.org/story/2018/07/30/justice-for-sexual-assault-survivors>

July 30:

From Wieffering:

Story 3: Liquor: Prospects for any victim are not great, but much worse if the victim was drinking or on drugs. Story also looks at MN's law on informed consent, which provides a loophole of sorts for assailants who've also been drinking. STAHL

Art: Joanna Howe (victim - in); Emily (?? - digital only?); MN statute doc (??)

Graphic: Yes (done)

Story 4: Training/turnover/ staffing levels for sex crime investigators. In Minneapolis, the sex crimes unit is "the penalty box," a place to dump cops with performance issues, or

investigators who've fallen out of favor with department leadership. But few departments have a dedicated sex crimes unit, and training/policies and procedures are weak or nonexistent at those that don't. BJORHUS

Art: Mike Martin (in); Paul Schnell (possible??)

Graphic:

Story 5: Prosecutors: Police forward only 25 percent of the cases they receive to county prosecutors, who in turn decline to bring charges in more than half the cases they receive. Victims are often left with little explanation as to why their assailants were never charged. STAHL

Art: Father hugging daughter (in)

Graphic:

Story 6: Sentences: Compiling sentencing data; one of the chief frustrations we hear from investigators and prosecutors is the light sentences rapists get, and how often judges depart downward. BJORHUS?

Art:

Graphic:

Story 7: We have a success story – a case that, against all seeming odds, went to trial and resulted in a conviction. Victim was drunk and doesn't remember all the details of that night; there were not witnesses, etc. Proof that if cops and prosecutors do their jobs diligently, "tough cases " can be won. BJORHUS?

Art: Possible victim (??); prosecutor (??)

Graphic:

Story 8: West Valley City, Utah, police department introduces new, trauma-informed training for sex crimes investigators. Dropouts decline; referrals to prosecutors and convictions climb, according to a before/after study conducted by Brigham Young University. STAHL

Art:

Graphic:

We've also been poking around at a story about the unusual relationship between victim advocacy groups and police, specifically whether advocates may contribute to the high number of women who "stop cooperating" on their cases, and if cops use the "victim first" philosophy to justify a passive approach to so many cases. But we are having a tough time making progress on this angle, which we've not seen tackled elsewhere.

July 19:

I've got 2 big findings from the regression analysis that you can start noodling.

The main thing to know before you look below is that the way this particular regression analysis works is that it allows you to put in several variables – key factors that influence the outcome of a case – and then see which of them have the biggest influence on that outcome. I tried running this with sent for prosecution as the outcome, and then separately as charges filed as the outcome. The charges filed one comes back with better results.

I tried many different factors and then winnowed out ones that have little or no impact. Ultimately I found that these factors were the strongest – victim involvement, whether it's stranger or acquaintance case, whether there was a delay in reporting and whether there was a named suspect. When you include multiple factors like this, then your results allow you to say that this finding holds true regardless of all these other factors.

Key findings:

1. The odds of charges being filed are 25 times greater if the victim cooperates throughout the case, even after accounting for other factors such as victim involvement, whether they have a named suspect and whether there was a delay in reporting the case.
1. The odds of charges being filed are nearly 4 times greater if the perpetrator was a stranger to the victim, even after accounting for other factors such as victim involvement, whether they have a named suspect and whether there was a delay in reporting the case. (Another way to think of this is that if they have two cases where victims are both cooperating, they have named suspects and there wasn't a delay – but one is a stranger case and the other is an acquaintance case – the stranger case is more likely to result in charges filed.)

Unfortunately, it's not giving me any significant findings regarding intoxicated victims or not. It's showing that the odds are just slightly lower that charges will be filed (or that it would be sent for prosecution) if the victim is intoxicated, but it's not enough to call that a good finding.

When I get back, I need to run my analysis and the wording on these past the professor who helped me. Perhaps she will catch something I missed, as well.

July 18:

I met today with Christina Knudson, a statistics professors in the mathematics department at St. Thomas University. She helped me run a logistic regression on the project data.

She recommended that I first exclude records where ALL the key variables we are using are coded as “unknown,” so I added that to my code and also excluded any records where the “sent for prosecution” field is unknown, since that is the dependent variable we are using. This ended up taking out about 30 records out of more than 1,000.

I also set up my code so that it creates new binary values (a 1 or a 0) for the variables I want to use. For example, it creates a new field called involve that is 1 if the victim cooperated throughout the process and a 0 if she did not. It’s NA if we don’t know.

I ran a first batch of models using “sent for prosecution” as the dependent variable. Next I want to try using “charges filed” and see if it comes up with anything radically different.

I asked Christina if I could run past her any wording that we plan to use in the stories and she said most definitely yes she could help with that.

Also, she asked if she could get a copy of our data when we’re done. She’d like to use it in her teaching.

July 11:

Linda Williams - senior research scientist; Justice and Gender Based Violence Research Initiative, Wellesly College

LA study - looked at how factors influence things; our study looking at things they didn’t do; We looked at was it cleared by arrest; also looked at some jurisdictional differences; we selected jurisdictions that use NIBRS so it had better data on the outcomes of the cases -- so you could see cleared by arrest and cleared by exceptional means (one-third of cases are cleared by exceptional means); we are trying to divide up victim drop out -- that victim actively dropped out or the cops sent letters or called and just can’t find them

Seeing African American women tending to drop out more -- don’t trust the justice system

Her study -- thinks the conviction rate is very similar (perhaps 6%)

We found delay in reporting was important in whether or not there was an arrest

Variables -- sometimes people combine some of these; “risky behavior” really predicts outcomes (this lumps in alcohol, drugs, walking alone at night, going to a bar); victim involvement; delay ; cases with physical evidence of force; a weapon was involved?;

Victim involvement -- increases likelihood of arrest 7 times (that’s our biggest findings)

Controlled for jurisdiction

Multi-collinearity - looking to see if they are so closely associated that they mean the same thing
We didn't have to exclude any variables from that

Victim not recalling the result -- not as big a factor as the others, but we think this is going to be important

Resources (officers/prosecutors per people in the jurisdiction) predicts exceptional clearance
--she is sending me a paper on that

Mental illness -- we noted if victim said she had a mental illness or was taking medication -- that really affected outcomes

Webinar that includes some of their findings:

https://www.youtube.com/watch?v=jA_mLuWfU5U

Less than 1 in 5 cases is cleared by arrest
Close to 30% were exceptionally cleared

Of the cases where charges were filed -- 53% resulted in a guilty verdict (189 cases) but they tended to be cases with a guilty plea (not a jury verdict)
Prosecutors told us that they were really doing a negotiated sentence, more than a negotiated plea.

Predictors of case outcomes:

Legal factors -- suspect physically assaulted victim; victim didn't recall details; weapon use; collateral injury; type of resistance (verbal, physical, both); reporting within one day; number of witnesses; victim willing to cooperate; physical evidence

Extra-legal factors - age; race; relationship to suspect; "risk-taking" behavior; questions about victim's character; mental illness or mental health issues; motive to lie

Non-strangers more likely to result in arrest

That was the only extra-legal factor in the LA study found to be significant; in our study we found several other that were significant

"Risk-taking" behavior were significantly less likely to end in arrest

If victim had mental health issues were significantly less likely to end in arrest

Victim not able to recall the assault -- were significantly less likely to end in arrest

June 26:

I realized that the requests we made to the agencies other than Minneapolis and St. Paul included an inordinate number of child or incest cases that we are weeding out. I asked if we had weeded a similar share of cases for the two big cities and Brandon told me that when he requested that data he asked them to provide only adult-victim cases. Well, we didn't do that for the other agencies. So we ended up with a different universe.

That puts a huge flaw in our statement that we have a 50% random sample of cases from these other agencies. So we're going to circle back and ask the agencies to attach more info to the original list of cases to identify which are child victims and incest victims -- then I can figure out what share of those cases we now have.

I discovered that Rochester had already done that and it turns out the adult cases we requested account for 50% of the adult cases. But we don't know if that's the case in the other jurisdictions because we don't know how many adult cases there were. I also discovered that Brainerd and Brooklyn Park gave us cases that were closed in 2015 or 2016 -- not the ones that were reported those two years (which is the universe we have for the others). So we're circling back to them and asking for revised lists of case numbers and we'll have to figure out how to proceed from there.

June 21:

When the Star Tribune first set out on this story, we thought it would be fairly simple to find out how many rape cases result in conviction.

So, we hunted down the available numbers for 2016, which showed:

- 2,321 rapes were reported to law enforcement
- 893 rape investigations were cleared.
- 611 sexual assault cases yielded convictions in the court system.

On the surface, it might appear that those three numbers go together and paint a picture of the outcomes of rape cases. But they don't.

"Cleared" doesn't necessarily mean someone was arrested and charged. Even more problematic is that the court's definition of a sexual assault case is broader than what law enforcement counts in their rape numbers. And then pile on the fact that some of the cleared cases or those that resulted in a conviction were actually reported in a previous year.

So, what happened to the 2,321 rapes reported in 2016?

Liquor cases analysis

<http://strib-data-internal.s3-us-west-1.amazonaws.com/projects/rape/liquor.html>

Key takeaways:

- About 30% of all the cases we reviewed had an intoxicated victim. That percentage holds steady across departments.
- The percent that are sent for prosecution is just slightly less when the victim is intoxicated, but it's not a big difference
- Most of that is probably due to the fact that the victim drop out rate is quite a bit higher (39% versus 29%)
- The ones that are sent for prosecution, though, are very likely to be declined. Just 9% have charges filed, compared to 15% of those where victim was not intoxicated.
- We also see some lower numbers in the police investigation work – less likely to collect all the evidence, interview all the witnesses (big gap on this one), slightly less likely to interview the victim, but suspect interview rate is the same. The witness one is the only one that really stands out.
- One interesting variation: When the case involves a suspect of color, the share that are arrested in cases involving an intoxicated victim isn't all that different than in cases where she wasn't. But.... When the suspect is white, 12% of suspects in intoxicated victim cases are arrested, compared to 32% in cases where she wasn't.

June 20:

To settle a nagging question in my brain, I re-ran our main analysis using a 50% random sample of Minneapolis and St. Paul cases, plus the other agencies (which were already a random sample). In other words, I based this analysis on a full "random sample" – not one that could possibly be skewed by Minneapolis and St. Paul being more complete than the other agencies.

The result is almost zero change. Some of the percentages budged by a point or two, but all the trends hold true.

Sent for prosecution:

Sample: 27%

Original analysis: 25%

Charges filed:

Sample: 12%

Original analysis 12%

Conviction:

Sample: 7%

Original: 7%

Victim not interviewed:

Sample: 34%

Original: 34%

Named suspect not interviewed:

Sample: 47%

Original: 49%

Not all evidence collected:

Sample: 58%

Original: 60%

Not all witnesses interviewed:

Sample: 46%

Original: 46%

Victim drop out rate:

Sample: 33%

Original: 34%

You can see all the results of the sample here:

<http://srib-data-public.s3-us-west-1.amazonaws.com/projects/rape/sample.html>

June 6:

Revised pub date for first story: July 22

Day 1 backout schedule

JUNE

04	Revised Day 1 to Eric
05/06	Rene and Suki feedback
07	Day 1 back to reporters
11	Day 1 final to Eric
11	Randy Lebedoff review
12/13	Final round of feedback/edits
14	Back to reporters for final edits/buff and polish
19	Final read by Eric, Suki and Rene
22	Day 1 moved to rim

Day 2 backout schedule

JUNE

08	Day 2, draft 1 to Eric
11/12	Rene and Suki feedback
12	Day 2 back to reporters
18	Revisions back to Eric
19	Rene and Suki feedback
20	Back to reporters
22	Final read by Rene, Suki and Eric

May 31:

We removed Wright County from the analysis because Jennifer found out that the only cases they sent us were ones that resulted in “no arrest.” They didn’t follow the public records law which requires them to provide us all closed cases, regardless of whether there was an arrest or not. They told us they have a “policy” that they only release the non-arrest cases and the county attorney releases the cases that had an arrest.

May 18th:

I finished the analysis looking for repeat offenders. Data is in Dropbox in the other data folder, file named: “rape_suspect_matches_May9”

The sheet in there called “Final repeats list ” has 50 people on it where we have found them in more than one case. There are various ways these guys got listed as “repeats”:

- This might be that our AirTable case mentioned another case they were a suspect in
- Our AirTable case might have mentioned that there was a second victim
- That we found prior court cases with criminal sexual conduct charge(s) that were either convicted, dismissed or acquitted
- That we have multiple cases where they are named in AirTable

For each person I’ve listed the total number of cases where they were named, the number of cases we have in AirTable with their name, and the number of court cases. Sometimes the court case matches their AirTable case. I’ve noted that in the notes. Prior cases are also noted.

I’ve also made a column for number of victims. This usually matches the number of cases unless I had evidence to show that there were multiple victims or there’s one guy with two cases involving the same victim. As a result, consider this field a “guess” on the number of victims.

There were two people (listed at the bottom) that you had on your paper list but I can’t find any reason why they were on there. AirTable doesn’t indicate any additional cases and I didn’t find them in court records – Cisco and Rabago.

Also there were three people that I don’t think you searched for in the vault (but we already know they have multiple cases): Trevon Ingram, John Hoffman, Devon Barclay. I’ll leave it up to you whether it’s worth running their names. We already know they are repeaters.

May 11th:

I built data highlights pages to share with sources. The main one is the top-level highlights:

<http://srib-data-public.s3-us-west-1.amazonaws.com/projects/rape/highlights.html>

Then there are others for Minneapolis, St Paul and St. Louis County that show how each agency stacks up to others on our key measures. It's easy enough to build a new page for another agency by taking the code and replacing any instances of the name (and I named dataframes with "stl" or "stp" or "mpls" that need to also be switched out)

Chase also set up a public bucket on the S3 server so the pages can be visible to the public.

May 9th:

I ran the suspect names against criminal sexual conduct cases from MNCIS through dedupe.io. It only turned up a few people that I didn't find in my SQL join and I think most of those are from the data cleaning Jennifer and I did after I ran the SQL join. Interestingly, it only found people who had a date of birth in our suspect list (about half our people don't have dates of birth-- either not listed in reports or were redacted).

The results of the join were mostly the court cases that match the cases in our police reports database (at least that's my best guess); but there were about 25 people that turned up with other cases that we don't have. Most were older cases or were in other counties, etc.

Next up is to run the names through MNCIS in the vault where we can turn up dismissed cases for these people.

May 8:

I grouped the cases into buckets:

One is simply "sent for prosecution" or "not sent for prosecution" (field called group2)

The other is "sent for prosecution", "assigned to investigators/not sent for prosecution" and "not assigned to investigators" (field called "group")

In the story I think we're going to use group2, and then focus the analysis on the "not sent for prosecution" cases to be able to show just how much police failed to take action on these cases.

Previously, we had been planning to use stats that represented the whole universe of cases, including the ones that were sent for prosecution and the "good" ones that resulted in conviction.

Here's a rewrite of the data points for story 1 using group2:

The Star Tribune examined nearly 800 sexual assault reports from police agencies across Minnesota and found that only one in four were sent to prosecutors.

Our review showed that the cases ***not sent to prosecutors*** sometimes were not even assigned to investigators, and even when they were, investigators often failed to take basic steps to try to solve the case and hold the suspect accountable.

- Investigators failed to interview the victim in more than 70 percent of the cases.
- Even when there was a named suspect, investigators failed to interview that person in 65 percent of the cases.
- In half the cases, investigators failed to interview potential witnesses.
- In about one-third of the cases that were reported immediately to police, investigators failed to collect all potential evidence, such as DNA or surveillance video.

We met today with Randy Lebedoff, our lawyer, to talk about hounding Minneapolis (and to a lesser extent St. Paul) for data requests that they haven't filled yet. We put in our request to Minneapolis for all these case files a full year ago and we're still waiting for some of the cases -- although we don't know exactly how many. We also think their pace has slowed and that they have started redacting more information than they are supposed to. And we have another request pending with them for policies/procedures/manuals related to sex crime investigations and we haven't gotten anything on that. In St. Paul, we have all the reports, but we are also waiting for the policies/manuals, etc request.

It sounds like the other agencies we're waiting for are starting to come through. Jennifer thinks we should have data rolling in over the next two or three weeks.

However that won't be fast enough for the first story. Rough draft is due May 23rd.

May 4:

I've been working on finding "repeat" offenders among the suspect names that we have. To this required first pulling the names out of AirTable and doing quite a bit of cleanup. In Airtable there are sometimes multiple suspects listed with the same case (all in one field, separated by semi-colons), so I flipped that out so there are separate rows for each suspect. Then Jennifer and I went through about 100 of them that didn't have dates of birth or middle names. We found quite a few dates or at least a middle name or initial, plus a few misspellings. (Turns out some of the agencies have been redacting the suspect's date of birth; other times it's just not listed although there might be an approximate age).

I first did a join in SQL, matching up with all cases (from both old and new court databases) that were identified as "sex crimes." In the old data we have some dismissed cases, but the newer data is only convictions. I found about 50 matches -- but in many cases it's likely just the conviction record for the case that we have in our database.

Suki agreed to pay to use dedupe to try for a better match. That turned up a few more matches on the first pass (matching on names and DOBs), but it ignored any of the suspects that we

didn't have a DOB for. I'm trying to see if I can do a second run with only name matching to see if I capture any others.

I also started working on graphic ideas. The main idea I'm fleshing out right now is a breakdown of what happens to these cases:

770 cases that we reviewed

- 25% are investigated and sent for prosecution
- 15% are not even assigned to an investigator and are never prosecuted
- 60% are assigned to investigators but never prosecuted

Of those 60% -- the majority are cases where the police failed to do 2 or more of the basic things we're tracking (interview victim, interview named suspect, visit scene if applicable, collect potential evidence, interview potential witnesses); Victims drop out in half of these cases; But interestingly, police had a named suspect in 75% of these cases.

The remaining 40% the police did almost all the basic work---and they even had a named suspect in about half of the cases; but they also lost about half of the victims;

May 1:

I added some new results to the data page today, focusing in on the issue of victims dropping out.

http://srib-data-internal.s3-us-west-1.amazonaws.com/projects/rape/rape2018.html#victim_involvement_rates

As we've discussed, victims have dropped out of about 35% of our cases and this seems to have an out-sized impact on whether cases get sent for prosecution.

The chances of a case going to prosecutors drops dramatically if the victim stops cooperating. Only **3%** of cases where victim dropped out were sent to prosecutors; **41%** of cases where victim cooperated were sent to prosecutors. Considering our overall rate is about 25%, that means this is a huge variable affecting that rate.

And, of course, as we discussed in today's meeting -- we have zero convictions in cases where the victim dropped out.

The data also shows a clear pattern that police do less work on cases where the victim drops out.

- Police interviewed a suspect in **15%** of cases where victim dropped out, compared to **57%** for those where victim cooperated. (cases with a named suspect)
- Police did not interview all potential witnesses in **55%** of cases where victim dropped out, compared to **43%** for those where victim cooperated. (only cases without a delay in reporting)

- Police did not collect all potential evidence in **33%** of cases where victim dropped out, compared to **26%** for those where victim cooperated (only cases without a delay)
- Investigators visited the crime scene **15%** of the time when victim dropped out, compared to **29%** of the time when victim cooperated (only applicable cases)

Finished this:

For Jennifer -- calculating conviction rates from the Texas data. Data is stored in Dropbox under the "other data" folder. Here's what she's looking for: Really, I'm looking for very basic numbers. I really don't even need them broken down by year. For the whole time period, I need:

the number of total offenses reported

the percentage of those that were charged

the percentage of all total offenses reported that resulted in conviction

the percent of all those that were charged, that resulted in conviction.

Note: conviction can be plea or trial.

April 30:

Built draft of rape analysis that reporters can share with sources:

<http://strib-data-internal.s3-us-west-1.amazonaws.com/projects/rape/highlights.html>

Decided to switch to run-and-gun approach rather than a 5-day publication in November. Aiming for publishing first story (a big one) on June 24th. Then roll out other stories on occasional basis, plus react to fallout and new victims coming forward.

April 18:

We got the discrepancy between Open Semantic and Air Table resolved. Hage also looked through the "not for database" cases to make sure none of them were incorrectly tagged.

April 17:

We've got a big discrepancy between the number of cases in Open Semantic and the number in AirTable. I think I may have gotten us a bit closer to figuring this mystery out....

- 1) There are 24 records tagged as "vulnerable adult" that are not tagged as having been reviewed by someone. I'm guessing that when you added the vulnerable tag, it lost the reviewed by tag (you have to put them all back on each time you edit this)
- 2) There is also another record that I found listed as not reviewed – LisaLeonAppleValleycase.

Here's the math on the Open Semantic side:

Not for database: 189 (including 148 that are not tagged as "reviewed by")

Reviewed (for database): 903

Needs review: 5

Missing review tags: 25

This gives me a grand total of 1,122 (it says there are 1,121 files in Open Semantic right now; not sure why I'm off by 1).

3) I think there might be cases in Open Semantic that haven't been reviewed or perhaps are tagged incorrectly. There are 388 files in Open Semantic tagged as "reviewed by Jennifer" and not tagged as not for database. So this SHOULD BE the total that shows up in AirTable. But there are only 346 records in AirTable with Jennifer's name on them. That's a discrepancy of 42 records.

I'm not implying that Jennifer did something wrong here. Maybe there's some weird tagging thing going on. Brandon told me yesterday that he found at least one case that was listed as "reviewed by Jennifer" but it hadn't truly been reviewed.

Interestingly, the cases that are listed as Reviewed by Brandon and Reviewed by Abby are the opposite – there are more cases in AirTable than I can find in Open Semantic (1 extra for Brandon and 2 extra for Abby). I suppose somebody put a wrong tag on in Open Semantic, which might account for a couple of the extras with Jennifer's name on them.

If we re-tag the ones listed above (under #1 and #2) and review the 5 in there now listed as "needs review", we would be just 30 records shy of having AirTable and Open Semantic match. So maybe the 42 records with Jennifer's name on them might resolve that – perhaps some were not reviewed; perhaps some need the "not for database" tag; perhaps some should be tagged as Brandon or Abby.

Worked on some data updates today:

Percentage of cases sent for prosecution that result in conviction:

About 31% of the cases sent for prosecution result in conviction. [Chart here.](#)

Suspect person of color:

Nearly 60% of our cases with a named suspect involve a person of color. [Chart here.](#)

In these cases....

-- about 33% are sent for prosecution. That compares to 32% sent when suspect is not a person of color. [Chart here.](#)

--about 19% have charges filed. That compares to 13% when suspect is not a person of color. [Chart here.](#)

--about 12% result in conviction. That compares to 7% when suspect is not a person of color. [Chart here.](#)

In other words, we have a very high share of suspects who are people of color. However, they don't appear to be treated too differently based on the data we have so far. The percentages are too close together (with possible exception on the conviction rate). However, it's worth keeping our eye on this as we continue to add cases.

Dave's questions:

The police index score is still very much a work in progress, but yes it appears that better police work yields more referrals and convictions. I need to spend more time on this to refine it and make sure we have solid results.

And yes, victim cooperation is clearly an enormously important element. But I think what I'm seeing in the police index score analysis is that even in cases with victim cooperation, sometimes the cops just don't do enough. I want to isolate in on those cases that don't have huge challenges for the police – victim is cooperating, she reported in a timely fashion and there's a named suspect. You would think the police would do more thorough work on those, but I'm finding that's not always the case.

April 11:

Jennifer and Brandon have been going back through cases that were already entered into the database and filling in fields that somehow were left blank. In many cases, these were fields that we added mid-way through the data entry. That has really helped give us better findings in the data.

I also started working on what I'm calling a "police index score". I'm giving them points depending on whether they did something or not. Here's the general logic:

- if suspect named, did they interview suspect
- if victim cooperated, did they interview victim
- if no delay, did they collect other evidence
- if no delay, did they collect all the potential evidence
- did they interview all potential witnesses
- if applicable, did they go to the scene
- was investigator assigned

The thing I'm having trouble with is what to do in cases where it's not applicable. For example, I can't give them zero points for not going to the scene when it wasn't applicable. Right now I'm giving them 2 points for a positive, 0 for a negative and 1 point for these weird non-applicable situations.

Ultimately I'd like to be able to say something like ... X percentage of the time, investigators failed to do hardly anything; X percentage of the time they did almost all of these things. (leaving kind of middle ground that might end up being mush that we ignore)

And then I'd like to be able to cross that "high" level of work versus "low" level of work against outcomes and types of cases (do they do less work in cases where victim is intoxicated and it's a non-white suspect? Do they do more work when it's a stranger case versus acquaintance?)

This would also make it possible for us to find cases that you would expect would have a "high" level of work --- no delay, victim involved, named suspect -- but the cops didn't do much of anything.

March 30:

Finished splitting Cass County and St Cloud files.

Worked on R analysis; discovered that they did enter some of the "open" cases from St. Paul into the database. Jennifer is going to delete them. But I also added a filter in my R script to winnow those out. That brought St. Paul's conviction rate up higher (from 3% to 5%); and it eliminated many of the records that had NULL values in key variables. Jennifer is going through the remainder of those (fields like conviction, suspectarrested, chargesfiled, etc) to see why we might have null values.

Next week I need to keep looking for those anomalies in the data and continue to hone the R analysis, which is now up on S3:

<http://s3-internal.s3-us-west-1.amazonaws.com/projects/rape/rape2018.html>

March 29:

Files arrived from St. Cloud and Cass County. However both sent all the cases lumped together in one (maybe two) big PDFs. So I'm splitting them apart. I finished the 2015 Cass County cases today.

Working on the R analysis; I've got it set up to import, do some cleanup and run some of the main queries. Need to do more on the cleanup part and then making sure we have all the tables and charts we want -- and in a good order, with the most important ones at the top perhaps.

March 27:

There's a file from Moorhead with pictures in it, making it too big for Open Semantic. We need to remove the pictures and store them separately.

Chase is swapping out the Open Semantic, upgrading to a bigger, more powerful server. Only change for reporters is a new URL.

I've started working on building an analysis in R that can be re-run each week. Need to set up an FTP solution for transferring the html file to the S3 server so it's more shareable.

March 18:

I've finished cross-checking. There are 63 records tagged as both "Not for database" and "Minneapolis" that I need you to go through and double-check that this is truly one that isn't meant for the database. I searched Airtable for all those case numbers (searched on a partial case number to cast a wider net) and didn't find any of them.

There are 19 records tagged as "needs review." This includes the one St. Paul one you caught on Friday, another that's labeled "AnnaAslani" and the remainder are from Brainerd.

March 16:

I am NEARLY finished cross-checking the AirTable database with Open Semantic and tagging all those cases you entered before we launched Open Semantic. I think I have roughly 50-ish cases to go through.

In doing this, I've seen a few things I wanted to alert you all to:

1. Please be super cautious entering the case numbers in AirTable. I found several that were either missing a digit or had an extra digit. Luckily I was able to hunt down the file and compare reporting dates and other details to make sure it was the right file. This is a crucial key for us to connect the data to the original documents. So it needs to be solid.
2. I've run across at least a half dozen cases that were reviewed by 2 people and are in AirTable twice. When I finish tagging, I'll turn my attention to this and we'll have to figure out what to keep, merge, ditch, etc.
3. There were 16 Minneapolis cases where the PDF didn't make it into Open Semantic. I ended up pulling them out of the batch files and uploading them. I may still encounter more because the 50 cases I have left to review are all Minneapolis cases. So this is a case in point in how this cross-referencing is paying off.

Right now the "needs review" bucket has 134 cases. I see that quite a few of them are from Brainerd (I'm assuming you're all still working on those?) and as I said above, roughly 50 are ones entered in AirTable that I haven't cross referenced yet. Any remainders will likely be those ones you chose to not put in the database and we'll flag those accordingly.

Problem with Mankato records: All 66 files are in Open Semantic – I matched against the S drive folder where Sue first put them. And all are marked in Open Semantic as having been reviewed. However, Open Semantic says 54 were reviewed AND entered into the database. Airtable only has 51 records for Mankato

Unfortunately to find them all in Open Semantic, you have to run 4 separate searches ---

“MK15” – 31 records (6 listed as not for database)

“MK16” -- 29 records (5 listed as not for database)

“MK2015” – 3 records (0 listed as not for database)

“MK2016” – 3 records (1 listed as not for database)

So we have a discrepancy of 3 records.

I see a few possible explanations – that there are 3 that weren’t entered into the database that either should have been or should have been tagged as “not for database” (but wasn’t). Or 3 got listed in AirTable as belonging to another jurisdiction.

If it’s the third option – assigned to the wrong department in AirTable – we’ll be able to figure that out once I finish the tagging. (I’ve got less than 200 records left)

To find the other options, we’d have to go through each one and check that it got entered everywhere and tagged properly. We can add that to the to-do list for a later date.

March 6:

Got Brandon, Jennifer and Abby going on Open Semantic. I’ve been tagging the documents that they’ve already reviewed by comparing across to AirTable. It’s slow going. Still working on the Minneapolis documents. There are some in Open Semantic that are not going into the database -- child abuse cases or cases that are still open -- that will be tagged as well. But going forward from here with new cases (such as Mankato cases that got added today), Brandon, Jennifer and Abby will tag them as “reviewed by...” on their way in.

I created directions for dealing with new files here:

https://www.dropbox.com/s/7gg0b37h4mvp4zk/RapeProject_OpenSemanticDirections.docx?dl=0

Feb 28:

Open Semantic is up and running

URL: <http://ec2-13-56-248-179.us-west-1.compute.amazonaws.com/search/>

Username: sriboss

Password: JenniferAndBrandonInvestigate

There are a few things Chase needs to tweak. It's also clear that we need to separately OCR the documents in the Dropbox file. Open Semantic doesn't put the text from its OCRing process back into the PDF; it just creates a separate text file on its site. So when you open the PDF itself, you can't search in the PDF if it's not OCR'd. Dan Barnes is looking into getting us Acrobat Pro, which can do batch OCRing on a directory.

Feb 26:

Dropbox for business account set up with myself as the admin and Brandon and Jennifer as users. The PDF reports that we have so far (just over 800) are being uploaded there. We are waiting to move the other project files (notes, court documents, etc) until Brandon and Jennifer can figure out a logical directory structure. Open Semantic is connected to the directory with the reports and has started OCR'ing the files. It's going slow, though.

Feb 16:

Met with Tony Wagner and Dan Barnes. Tony is assigning someone to set up laptops with 1tb hard drives for Jen and Brandon. Work will begin Monday. Dan Barnes will transfer over software and that kind of transition. He's also setting up the Dropbox account. The Open Semantic Search part won't be able to be connected until week of Feb 26th due to Chase being out of the office next week.

Feb 15 :

Afternoon meeting:

We went through the fields in the database and debated some that were questionable. Ultimately we added some, removed some and slightly changed up the language on some. The main ones we struggled with had to do with our attempts to document whether police went beyond the basics (basic interviews, assigning investigators, etc) -- like did investigators go to the scene, did they make an extra effort to contact reluctant victims, did they make extra effort to find a suspect, did they interview all potential witnesses, did they collect all potential evidence. These things are hard for us to nail down sometimes because we are dependent on what is documented in the report. For example, if there's a comment in the report that there might be surveillance video from a nearby store and no evidence that it was collected, we could say "no" the police did not collect all potential evidence. But if there's no mention of that video footage, then we can't answer that question. Ultimately we settled on the fact that these fields might not be super reliable for strong statistical measures. It really depends what we find. But at minimum, I think we'll be able to paint broad brushstrokes about the infrequency of police going beyond the basic things or not taking action unless witnesses, evidence and cooperative victims are handed to them.

Notes from morning meeting:

Jennifer is finding that different places are measuring “forcible rape” differently; some places are including founded or unfounded; some are including penetration or not penetration -- others are only including penetration cases.

Our universe has all -- unfounded cases, groping, not penetration, molestation, etc. We’re not measuring child victim cases. Could best be described as “sexual assault, rape, and sexual battery (no penetration)”

“Sexual assault” or “rape” -- to refer to any type of penetration (fingers, penis, oral); everything else they refer to as “sexual battery” (touching, groping, etc). Not clear if this includes child cases.

We need to understand the universe of cases we’re looking at and how we want to define our conviction rate based on that. Reality is that it’s not going to match any other jurisdiction because they are all doing it differently

Canada does a good job tracking stats nationally; Jennifer is going to see what universe they use and see if that’s something we can mirror. She’s also going to look at some other studies and see if there’s something we want to match

From Jennifer (Feb 15) via email:

I just got off the phone with Linda Williams, senior researcher at the Wellesley Centers for Women at Wellesley College. She is copying the methodology that the researchers used in that textbook I showed you at the meeting this morning, which studied the Los Angeles PD and Sheriff’s Office. With an NIJ grant, Williams is studying sexual assault cases from **six unnamed police departments** around the country (none in Minnesota), including a few large ones. They will have preliminary results in March that they can share with us. But she gave me the basics today:

1. Their universe of police reports is rape, sexual assault and “forcible fondling” (which the textbook called “sexual battery”). I like the term forcible fondling better than sexual battery. This does NOT include cases of flashing women, for instance. There must at least be unwanted touching.
2. They include any case where the victim is 13 years or older at the time of the incident, and it was NOT a case of incest.
3. They include victims who are men. (She doesn’t think the LA study included male victims. I will check.)
4. They include cases where the suspect may not be an adult. However, since there is no good information on the outcomes of juvenile cases, the final conviction number will not include those convictions. She realizes this is imperfect.
5. In all, one in 5 reports result in an arrest.
6. 10 percent of the total universe of reports are unfounded.

7. The victim does not cooperate (very broad term) in roughly one-third of the cases. But some cases where the victim drops out do result in convictions.
8. **Of all the reports of sex assault, including the ones that were unfounded, the conviction rate is a “pretty dismal” 6.5 percent. Strip out the unfounded cases and it rises to 7 percent.**
9. About 30 percent of the whole universe of reports are cleared by “exceptional means.” She is unsure whether we have this in MN – it’s a way police and prosecutors basically get rid of cases where there is an identified perp, but the victim won’t cooperate or they just feel the case ain’t going to work. Do we use this here??
10. Of those cases where there was an arrest, 72 percent had charges file.
11. Of those where charges were filed, 53 percent resulted in a conviction.

BACKGROUND NOTES:

Technical:

The reporters are reading each report and filling in fields in a form that we created using Airtable. We’ll be able to export that data to other software once we’re finished. In the meantime, it is possible to easily search and/or filter the dataset to go back and edit existing records. That’s something that isn’t available in Google Forms.

This link will take you to the database:

<https://airtable.com/invite/l?inviteId=invmzpkOX8HCbU2w6&inviteToken=c85a96bb996c9460e95be084a194b8a4>

This link will take you to the form for entering data:

<https://airtable.com/shrjBETubWRjYdyap>

Dropbox: We are setting up a paid Dropbox account that we’ll jointly use for all of the documents, data and reporting materials for this project. Jennifer and Brandon will be getting larger hard drives so that they can sync the dropbox files to their hard drive for easier accessibility.

Open Semantic Search: The data team set up Open Semantic Search on our Amazon server. This tool will OCR the documents we’re getting and allow reporters to search, tag and make

notes on the documents. This will point to the Dropbox folder containing the original files, with each report stored as its own PDF.

Sharing the documents with panel of experts: We have a panel of about a dozen experts who have agreed to review some of the cases. The plan is that we will set up a Google Doc for each expert and put links to the documents (in Dropbox) that we want them to review. They can then put their notes in the Google doc, right below the link.

Scanning of paper documents: We got paper documents from St. Paul (others so far have been via PDF), so we've enlisted Susan Hilliard (newsroom clerk) to use our portable scanner to scan the documents and store them in the S drive. One of us will have to transfer them to the Dropbox account.

How we devised the database questions:

There were several iterations of the database. Initially Brandon just had an Excel sheet where he was logging key pieces of information about the U of M cases that he went through first. We used that as a starting point, then started thinking about the ultimate questions we wanted to be able to answer:

Why are these cases so rarely charged? Why are rapists so rarely held accountable?
What happens when the victim cooperates -- does that improve likelihood of charges? When victim doesn't cooperate, then why?
How often do police attempt to interview the suspect? And the witnesses?
What happens when the case is intoxication/incapacitation vs threat or fear?
If a report was more thorough, did it increase the percent of cases were prosecuted?
The quality of the documentation -- are they including things that help corroborate force/fear, are they summarizing things -- are they producing a good package to give to the prosecutor?
What happens when they take the steps needed to investigate a case?

Then we started looking at "best practices" documents created by experts, including theses:
<http://www.theiacp.org/Portals/0/documents/pdfs/Communications/IACP%20Sexual%20Assault%20Response%20Policy%20and%20Training%20Content%20Guidelines.2017.3.23.pdf>

<http://www.evawintl.org/library/DocumentLibraryHandler.ashx?id=43>

<http://www.evawintl.org/WebinarArchive.aspx>

<http://www.theiacp.org/Police-Response-to-Violence-Against-Women>

<http://www.theiacp.org/Portals/0/documents/pdfs/ResponsetoSexualAssaultChecklist.pdf>

We also had email and phone conversations with some of the experts that have helped create these best practices and guidelines.

We determined that there are some key “buckets” we’d want to assign cases into -- such as, whether the victim cooperated throughout the case or withdrew, whether this was an acquaintance or stranger situation. So we knew we needed yes/no fields for those.

We spent a lot of time talking about which things we could discern from reports and which we could not. For example, the best practices guidelines talk a lot about “did the police do everything they possibly could?” We can’t really answer that question, posed quite that way. Because we only know what they wrote down. So, we can see for example that perhaps they wrote down the name of a witness, but there’s no documentation that they interviewed that witness. So then we can say they didn’t interview the witness. But we wouldn’t know if there are other witnesses not named in the report that they didn’t talk to.

There are essentially five types of things being collected in the database:

1. Basic info – dates, case numbers, disposition of case, etc
2. Fields that help put cases into groups – whether it was acquaintance or stranger, whether victim withdrew cooperation, whether there was a delay between incident and when reported to police, were investigators assigned, was it prosecuted, was there was a conviction.
3. Fields that document basic policing – did they interview suspect? Forensic interview with victim? Rape kit done? Other evidence collected? Was there something obvious that they didn’t do (like visit scene of incident, fail to interview named witnesses, etc)
4. Fields that document quality of the police report – did they include testing results? Do they document key things needed for prosecution (such as evidence of use of force or incapacitation)?
5. Fields for reporters to use for notes or to flag key cases

Data from the BCA/FBI:

From the BCA, we requested the number of reported, actual, cleared and unfounded cases reported by each police jurisdiction between 2010 and 2016 (broken down by year) for each of the Part 1 crimes. (Reported is the total number reported, actual is everything except unfounded cases). We’re going to use this to determine how often, across the years, rape cases

result in nothing (not cleared, not even deemed unfounded) and how that compares to other major crimes. Request made with Jill Oliveria on Nov 7.

--results are in the R file, rapeanalysis.Rmd

C:\Users\webstmj\OneDrive - Star Tribune\CriminalJustice\Rape

Jennifer also asked the FBI for clearance rates at the national level. Here's their email:

Please tell Ms. Bjorhus that the FBI Uniform Crime Reporting Program does not collect data for rapes that were "cleared for prosecution." Instead, the Program collects clearances based on arrests and exceptional means. Table 27 in Crime in the United States (CIUS) 2015 has the data she is looking for.

Below is a link:

<https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015/tables/table-27>

The same table is in CIUS 2016, but as Table 17. Here is a link:

<https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/tables/table-17>

Also, the unfounded data for rape that she is interested in must be ordered. We will email as soon as it is received.

I calculated clearance rates for major crimes and compared them to Minnesota.

It shows that Minnesota's 38% clearance rate for rape cases is on par with the national rate (37%). Of course, clearance doesn't necessarily mean conviction. They can clear cases by arresting someone or through an "exceptional means" clause and we don't know how many were cleared by an arrest versus exceptional means. And we don't know if those arrests resulted in charges.

Compared to other crimes, rape cases in Minnesota are cleared less often than murder (74%), human trafficking (81%) and aggravated assault (61%), but more often than burglary (16%), motor vehicle theft (18%), robbery (24%) and larceny (31%). That pattern holds at the national level too.

Minneapolis records:

Caresa Meuwissen in the records office is fulfilling Brandon's requests for all closed rape cases in 2015 and 2016. Brandon and I met with her and her boss, Mary Zenzen, at the beginning of November to discuss our request. They clarified that it was going to take a long time – and cost us a lot of money (more than \$6,000) – because the "closure code" in the CAPRS database isn't always correct. They don't feel they can rely on it to determine which cases are closed and which are not. So Caresa is opening up each case file and looking for additional clues and in some cases pulling up court records. We told them that we don't think they are legally allowed to charge us for that time. (My initial thought was that this was verifying the accuracy of the data; but Shiffer later reminded me that this could also be considered the part of the process where they determine whether a record is public or not. In either case, they are not allowed to charge us for that time). Mary was going to take that to her supervisor and get back to us. We

told her to proceed with fulfilling our request regardless and that, if necessary, we will pay the bill but continue to fight the charges. (updated Nov 9). They also told us that they were getting slowed down because some highly sensitive and/or high-profile rape cases were locked with passwords (the portion of the report that is only public after the case is closed). Caresa said she has to reach out to the investigative unit and try to hunt down password; many times, the password wasn't documented and they have to get IT to unlock the case.

We consulted our lawyer on this and also looped in city attorney Susan Segal. She quickly got involved and told us the city will not charge for the time spent determining whether a case is open or closed, but then she also said she didn't think records staff SHOULD be doing that. She said it's not their responsibility to determine if a case is open or closed -- they should just give us cases that are labeled in the database as closed. However, that made us worry that we would miss hundreds of cases that we should get because they know that the database coding is so unreliable. Susan and Mary Zenzen eventually replied that they will continue doing this work to ensure we are getting all the records we should, but won't charge for the time spent determining public vs private status.

As of early February, they are sending us files on a rolling basis -- once a week -- and we think we've gotten about half the cases, so far.

St Paul records

In August 2017, we also requested all closed rape cases for 2015 and 2016 from St. Paul police. They initially told us it would cost \$200, then later they revised that estimate to \$1,536. This was supposed to be for 6,147 pages (we don't know how many reports that would be).

Brandon and I went to their office and chatted with them about this in November. They showed us one box of documents (printed out) and they told us the other box was almost ready.

They asked what we planned to do with the data and Brandon was honest and upfront with them. Suddenly the following week, they said ["Given the sensitivity of the data we are releasing, we have decided to take another run through to make sure its properly redacted. This will delay the release for a little while, but we will notify you when we are done."](#)

We got the boxes in early February, with fewer records than they had originally told us, and total cost of \$950

The police department printed out the reports, put tape over any names that needed to be redacted, and then photocopied them.

Other agencies:

We identified 18 other agencies that account for large numbers of rape cases and we initially sent them data practices requests asking for the case number, date opened and date closed for all rape cases closed in 2015 and 2016. It turns out that some agencies -- like Coon Rapids and Fridley -- have these cases handled by Anoka County. They all complied with that (however some of it came on PDF; some of it in Excel; some of it just listed in an email). I converted all of those to Excel and we created a random sample from each.

We decided to capture 50% of cases from each agency -- except three agencies that had less than 50 reports each; in those cases we asked for all of them. We decided to go with a rather large percentage of cases on the assumption that some of the cases we request full documents on might not pan out -- either because they are child abuse cases or because the agency suddenly discovers the case is still open or something like that.

For each agency, I used RANDBETWEEN to assign a random number to each case number and then sorted it smallest to largest and took the top 50% of cases. The files where I did this are saved in the master Rape Project directory on Dropbox. Then we generated Excel files with the listed cases and sent those to each agency asking them to provide the full reports for those cases.

Anoka County, which is the largest of the bunch, balked at our request, saying it would take until at least end of summer and cost us "thousands of dollars." We knocked down our request to 40% of the cases and that eased their concerns, along with some kind words from Jennifer. They are also going to provide the records on a rolling basis, as they are ready. Hopefully we'll have all of them by early summer.

The other agencies are working on their requests, as of mid-February.

We ran into several situations where the agency told us they would charge us 25 cents per page for the first 100 pages, then switch to an hourly rate for the remaining. We checked with Stacie Christensen at the Data Practices Office and she advised that if the total request is over 100 pages, they should just start with the hourly rate and only charge that.

Other rape data:

<https://www.revealnews.org/blog/here-are-three-startling-new-stats-on-rape/>

<https://www.bjs.gov/content/pub/pdf/rsavcaf9513.pdf>

*On RAINN's website, we have tried to select the most reliable source of statistics for each topic. The primary data source we use is the [National Crime Victimization Survey \(NCVS\)](#), which is an annual study conducted by the Justice Department. To conduct NCVS, researchers interview tens of thousands of Americans each year to learn about crimes that they've experienced. Based on those interviews, the study provides estimates of the total number of crimes, including those that were not reported to police. **While NCVS has a number of limitations (most importantly, children under age 12 are not included), overall, it is the most reliable source of crime statistics in the U.S.***

Similar stories:

<https://www.theglobeandmail.com/news/investigations/unfounded-sexual-assault-canada-main/article33891309/> Methodology:

<https://www.theglobeandmail.com/news/investigations/unfounded-data-methodology-sexual-assault-claims/article33891819/>

The Belleville, Ill., News Democrat published a project called "A Violation Of Trust."

<http://www.bnd.com/news/special-reports/violation-of-trust/>

<https://injusticetoday.com/memphis-police-failed-to-properly-4d7685f87a5>

http://www.oregonlive.com/clackamascounty/index.ssf/2017/12/she_reported_a_sexual_assault_1.html

<https://projects.fivethirtyeight.com/sexual-assault-victims/>

In 1999, the Philadelphia Inquirer revealed that for nearly two decades, city police had been burying rape and sexual assault cases by marking a high percentage of cases as “unfounded” or by classifying them in non-crime codes. In the early 1980s, the police department rejected over half of all rape complaints as “unfounded”—an FBI category for cases in which investigation shows no crime occurred—at a rate five times the national average. The department stopped classifying cases as “unfounded” in such high.... Etc.

<http://inquirer.philly.com/packages/crime/html/sch101799.asp>

Database fields:

Who created this entry - Name of reporter who made the initial entry

Source PDF - Name of PDF report is stored in

Date entered -- Date that reporter entered this information initially

Department -- which agency handled the case (i.e. Minneapolis, St Paul, Anoka County)

Case number -- police case number

Incident date - mm/dd/yyyy when incident occurred

Were investigators assigned? -- yes/no/unknown

Investigator name(s) ---

Reported date -- Date incident was reported to police

Reporting delay -- Was there a delay between incident and when reported? yes/no

Suspect name(s) -- name(s) identified in the report

Suspect DOB -- date of birth of first suspect named in report

Suspect_alternateDOB -- use this field if the report lists a second possible DOB

Suspect a person of color? -- yes/no/unknown - is the suspect a person of color

College Students Involved -- are college students involved (either victim or suspect)? Choose yes for any college, not just the U of M.

Is this a statutory case? --

Accusations -- Description of accusations (as summarized by the reporter)

Case Type -- stranger/acquaintance

Legal type -- force or incapacitation (which could include asleep, drugged, drunk, mental)

Victim involved? Was the victim involved throughout the process

Victim interviewed - Did police conduct a forensic interview with victim?

Interview conducted at police station? yes/no/unknown

Was suspect(s) arrested? yes/no/unknown

Was case sent for prosecution? yes/no

If sent for prosecution, which county?

Charges filed? yes/no/unknown (this needs to be gathered from court records)

Conviction? yes/no/need to check/case still pending (this needs to be gathered from court records)

Sentence -- if there is a conviction (we should be able to get this from court data)

Victim male? - check only if the victim is male

Was victim intoxicated? yes/no/unknown

Did intoxication affect victims memory? yes/no/unknown

Suspect interviewed -- was at least one suspect(s) interviewed? yes/no/unknown

Did responding officers talk to any witnesses? (either eye-witness or friends of victim or suspect)

Did investigators fail to interview witnesses listed in the report? yes/no/not applicable

Was rape exam done? yes/no/unknown/not applicable/victim refused

Kit destroyed? yes/no/unknown/NA

Was other physical evidence collected? yes/no

Were testing results documented in report? Say yes if the report includes the results of any testing, such as rape kit or testing of other physical evidence

Is there documentation of incapacitation? For cases of that legal type. Only say yes if documentation is from a source other than the victim's statements (list items below)

Is there documentation of force? For cases of that legal type. Only say yes if documentation is from a source other than the victim's statements (list items below)

Incapacitation examples--

Force examples --

Language of non-consent used throughout --

Is there something that police didn't do? - yes/no (other than not interviewing). Examples: visit scene, collect surveillance video, not getting rape kit tested

Describe any other actions police could have taken but did not--

Closure code --

Date closed -- date that case was closed mm/dd/yyyy

Highlights and notes of the case -- these are notes for the reporter's use

People to interview -- notes for the reporter to flag names of people in this case they want to interview.

FBI definitions of sex offenses for statistical purposes.

Forcible – Any sexual act directed against another person, forcibly and/or against that person's will; or not forcibly or against the person's will where the victim is incapable of giving consent.

1. **Forcible Rape** – The carnal knowledge of a person, forcibly and/or against that person's will; or not forcibly or against the person's will where the victim is incapable of giving consent because of his/her temporary or permanent mental or physical incapacity (or because of his/her youth).
2. **Forcible Sodomy** – Oral or anal sexual intercourse with another person, forcibly and/or against that person's will; or not forcibly against the person's will where the victim is incapable of giving consent because of his/her youth or because of his/her temporary or permanent mental or physical incapacity.

3. **Sexual Assault With An Object** – The use of an object or instrument to unlawfully penetrate, however slightly, the genital or anal opening of the body of another person, forcibly and/or against that person's will; or not forcibly or against this person's will where the victim is incapable of giving consent because of his/her youth or because of his/her temporary or permanent mental or physical incapacity.

4. **Forcible Fondling** – The touching of the private body parts of another person for the purpose of sexual gratification, forcibly and/or against that person's will; or, not forcibly or against the person's will where the victim is incapable of giving consent because of his/her youth or because of his/her temporary or permanent mental incapacity.

Non-Forcible – Unlawful, non-forcible sexual intercourse.

1. **Incest**–Non-forcible sexual intercourse between persons who are related to each other within the degrees wherein marriage is prohibited by law.

2. **Statutory Rape**–Non-forcible sexual intercourse with a person who is under the statutory age of consent.