Jacob Interview

Tue, May 02, 2023 2:27PM • 25:24

**SUMMARY KEYWORDS**

reporting, crime, agencies, data, estimates, fbi, estimation, new york state, state, public, system, robbery, arson, compiling, specific, release, neighbors, new york city, question, based

00:00

Five years ago now.

00:02

And at that time, the FBI had a website to, like show crime data.

00:09

But it only had data through, I think 2014.

00:14

And it didn't actually have any graphs or anything, it was just a table showing agency level data for some agencies for index crimes

00:23

from like, 1960, to 2014.

00:27

So like, they had a lot more data available that they released, just like the master files. So I thought, like, a lot easier for me to understand data if it's visualized, and like a graph.

00:39

So I decided to just build that tool, just like a look at the data myself and try to understand it. And then since that's like, being able to see data without having a lot of like programming skills is really, really useful. So I made it public, just so people don't need to learn how to program or deal with the data themselves, to be able to understand the crop data. I see. And then I added in the other FBI datasets, I like the arrest data about hate crime data. So is there anything that surprised you when you make use of the data set by the FBI? Is there anything surprising or interesting findings that you can share with us?

01:24

Yeah, so the data they release, like the master files is all at all agency level. They don't do any kind of imputation from this, I guess. So it's basically like as raw data as you get. And like looking at, you could see like, there's clear, like data issues like arson, there's data on arson, and there's like a month with I think, like, 7 million vehicle arsons in one little city. So plainly, it's just a data entry error. But it's like something that you I didn't catch when they're producing this data, or releasing the data. And then you could just see some agencies, like Chicago just decided to stop reporting rate for about, I think about a decade. So just being able to visualize it, it's like super easy to find these huge data errors like that. And like New York City,

02:16

starting in about 2005, for about a decade, they stopped reporting monthly data, and they just reported quarterly data. So if you're trying to understand, like monthly credit, you're going to see zero for like two months, a year, three times as much as you'd expect in that third month, and then to for the next two months, and then triple the amount, the like the six months a year because it's quarterly.

02:41

I see. We found there, like, like you said, about arsons, you have very specific categories of the type of crime that you you're the the crime data tool, but as far as we know, the FBI only have eight types of crime that they they don't have that specific categories. Is that correct? So they only saw the index crimes are like the seventh or eighth if you include arson. Yeah. They consider like the most serious and best reported crimes. And that's what they report on. They're like official website. And then they're like annual crime, the US report. But if you look at the raw data, they actually do have a lot more crimes. I see. So for like, robbery, that's an index crime. So you have robbery count, they have that on their own site. But in their data they report robbery with with a gun robbery with a knife. And robbery, I think without any weapon. Looks like what they're showing is only a small amount of what they actually release in there, like herkes data. And that's like another reason I wanted to build the tool, because it's like they have this data, but they're not just making it easy for people to use. I see. And for the raw data. Did you go to the FBI website? Or did you go to like each agency and county levels data?

04:03

Through the FBI? Oh, I see. Okay. Um, and in our like previous email communication, you said you will give a talk this fall on the FBI is missing crime data and with a focus in New York State. And because our project also like specifically focused on the datasets that are made public by the New York, so could you just talk more, or maybe just share a little bit about your talk? And what's unique about New York State in their transition to the library system?

04:38

Yeah, so the talk is like broadly about, like, how much missingness there is and what the steps are to basically estimate the messiness and like the issues with that. And like New York State is probably the most perfect example of like, estimating missingness but doing a really poor job about it. So like the example where who I found where they

05:00

Your FBI reported, like a tiny rate of crime of violent crime relative to what New York State reported.

05:08

And that's because most agencies in New York state didn't report any data to the FBI, but they didn't report it to the state because the state accepted the old data system.

05:19

But then you have places like New York City, which has like 40%, about of the entire state's population. And it's by far the biggest city in the country. They didn't report and he did it to the FBI. But then the FBI still estimated the state level data, which like, you can't estimate a state level data when you're missing so much data. And especially this is like the biggest city. And you could see how bad that happened, because their estimates like

05:44

they could take a quarter or a third of the actual rate of violent crime that New York State reported. So it's just massively off of what reality is. So as far as you know, do you know like, what has prevented the New York, like YPD, or other agencies in New York City from making transitions and reporting the data efficiently? Yeah, so the conversion from the old summary reporting system to the newer NIBRS is collecting a lot more information, which means the agencies have to buy an entirely new data management system, which is pretty expensive. And then they have to retrain everyone who's inputting data, which is usually just the loan officers themselves. So it's a lot of money, just buying the system, and then a lot of time and my self retraining the officers. So I think agencies are sometimes reluctant to do that, especially if the state isn't requiring them to do that. And I also think, so what FBI said, I think, in 2015, that in 2021, or in 2020, that's the last year you report to the old system. And I think a lot of agencies just didn't believe them. Like, they probably thought that the FBI would give them an extension. So there's like nobody to really rush. actually kind of see that because

07:11

it like a couple of years before 2021. Like in 2018 2019 2020. There's a huge spike in like the agencies that start reporting. So I think they like kind of got desperate and realize that they need to start reporting. So I want to like, started doing a conversion. Yeah, we're still missing, like 30 or 40% of the country.

07:32

For agencies reporting something, it's just, it's a lot of money that they don't have or don't want to spend

07:38

a lot of time to do the conversion. And they probably just didn't expect the deadline to be real.

07:44

I see. And could you talk more about the like, draw the comparison between the libraries, and the previous the summary reporting system was the biggest difference between these two? And does the new system actually make the public easier or harder to understand a crime?

08:03

Yeah, so the big difference is a summary reporting system is its summary. So it's like aggregate data. So for each agency, you get just the monthly number of crimes and arrests.

08:16

For like a number of like a smaller number of crimes and neighbors. So if you want to know, like the number of murders that happened in New York City in a month, like you have that of the old system, you just don't really have anything detailed. There are some exceptions, homicide actually has more specific info, but for other crimes, it's just like the monthly count. And then neighbors, it's the National Incident based reporting system. So the incident part is really key. We have pretty detailed information on every reported incident of crime, including demographics of the victim, the offender, some incident characteristics, like where it happened, just like to category like school or sidewalk, it's not a precise location. And then you also have time of day, you actually know exactly when each person is arrested, which is new. So it's a lot more detailed information. And you also have more crimes reported. So like, you could start tracking fraud. You could track some sexual abuse that wasn't checked before.

09:23

That's just more detailed, more crimes.

09:27

As far as like the public understanding it, I think that's a much harder question. Because, like, yeah, we have more information, but it still has to be compiled in a pretty easy way for the public to understand. Just because, like it's complex, like, if you're talking about robbery, now, you could say like, Okay, I'm just gonna look at robbery, or robbery of women of a certain age, or, like, you could look pretty deep into the data. And I think that that's useful for in some ways you could

10:00

Seems like who's been victimized. But then it's more complicated because, you know, there's way more categories of crime to go look at. And is FBI responsible for compiling all those data?

10:12

Yeah, so I think at the federal level, it's completely voluntary for agencies to report data.

10:20

But the state level, some states, either in their legislature, they have a law saying, agencies need to report it, or some like the post certification, which speaks basically accreditation for being a police department, you have to report it.

10:37

So I feel I can't mandate agencies report. But once they do report it to the FBI, they're responsible for

10:43

basically checking to make sure it's correct. And then compiling all of like the master files and all the reports they make, and then sending it to the public. I see.

10:54

And so, as we know, like, at this point, a lot of states have stopped submitting information through the summary reporting system, but they haven't made the full transition to the libraries. So for those dates, is there any possible way for us to gather a full picture of the crime in those states or counties? Do you have a go to each agency's website and collect data? Or is there any other way?

11:19

Yeah, so some states will collect it themselves.

11:24

I think New York is one of those, California is one of those. So an agency will report data to a state, another state will report it to the FBI. So even in states where they're not reporting Magirus data like a California, an agency will report flow state, and you can go to the state website and download the data, though it's not always as detailed as what eventually goes to the FBI, because there's some like filtering before they release it to the public.

11:52

But otherwise, yeah, you'd have to go to like each individual agency. But I'd definitely start at the state level. Does the state have like, is there a state ECR system where it has the same kind of data available at the state level for those agencies? And if not the government agencies? Okay. And another question is, do you have any knowledge about methodology and techniques that the FBI used to came up with like to come up with those estimations? Do they make it public? Or do you have any insider knowledge about their methodology? So FBI, for their estimation, for neighbors, they worked with the Bureau of Justice Statistics, and then Research Triangle Institute International,

12:41

which is like a pretty big criminal justice thing thick. And they do have some like PDFs online detailing how they did their estimation. It's not super specific. So they'll say like, if an agency is missing, we'll replace the data with like a comparable agency, though they don't like specify super clearly what a comparable agency means. And as far as I know, they also haven't released anything about like, their steps to verify that their estimation is correct. And from what we've seen in New York, your estimation is at least in some cases, very not correct.

13:21

Have you tried any way to verify their mess estimations or

13:28

their results

13:31

compared to some of their data from like, past releases.

13:38

For 2019 data, they have the old system. So it's like almost perfect reporting. It's like 90 or 95% reporting. And I've compared that with their 2020 estimates from nybridge data.

13:50

And then 2020 release, they also the estimated 2019 diapers, they use that as a comparison here, which is really weird. I like in a lot of those cases, the numbers don't align with what they reported initially in 2019, even though it should be if their estimation is correct.

14:08

And I know, Jeff Asher, who's like a data analyst in New Orleans has also done the same thing. I think he's done more than I have on this front. Yeah, we talked to him last week.

14:21

He's great. Yeah.

14:24

And the next question is, do you know, how much do you know about the accuracy of the estimates in different types of crime?

14:39

So I really looked like compare the different types of crime they've reported. Yeah, sure. I just I know from the National Crime Victimization Survey, there's like, huge differences in

14:51

reporting to the police on crime. So if you're looking at like the true rate of crime, that's definitely different from estimation. And the fact that you're estimating based on unknown

15:00

Under random sample of agencies, for crimes that are themselves underreported, it's gonna be like, far off from reality. Like the true number, though, it's always gonna be like impossible to get the true number. I'm not sure how much the estimation differs based on the crime itself. Yeah, because we also found that, for example, for the New York State, the FBI only give estimates for certain types of crime, for example, like rape, bog array or,

15:30

and they leave it blank for murder and robbery. So, and we're trying to figure out the reasons behind why they give estimates for some types of crime, and they leave it blank for others. So do you have any knowledge of that? So they say in one of their reports, about their estimation procedure, that they have a process to determine when the estimate is like, too wide. So like, if they think it's too far off? And they won't report it? They'll they're not super specific on that. And also, like, in the case of New York State, like they shouldn't be reporting anything just because they're missing so many other agencies.

16:10

And it's weird that they're reporting some crimes, not others, my guess is that murder and robbery are relatively rare, at least compared to other crimes, that might be their concern that they're too rare to estimate properly. And therefore, they're not reporting the rest of it. I see.

16:29

And is there any way for us to know,

16:32

the FBI method evaluating evaluating the bias or accuracy of their estimates?

16:40

I would check their documentation and see if they have anything more specific, or just like FOIA request for actual method, since that should be public. Okay. Yeah, yeah, definitely. Um, and we are also trying to figure out the impacts and consequences of those missing missing reports and inaccurate estimates. So is there an example of those data being quoted or used by agencies or politicians? Or what's your thoughts on the consequence of the FBI is missing crime data?

17:18

I actually have like a whole blog post on how the media messes up, like solution data. Do you share it with me? Yeah.

17:28

I try not to follow like crime news, just because it annoys me.

17:33

So I don't really know if any politicians have said anything. But when the 2020 crime data came out, or I guess the 2021, crime data came out, like pretty much every news agency reported on the numbers. And they reported like,

17:49

as if there was a true number, and then later on, in the article, they're all like, these are estimates and like, experts say you shouldn't trust them. But like, I think for most people, if you're gonna read like the headline in the first paragraph, you're like, Well, this is the true number of crime in my state or in the country.

18:08

So I think people are going to be making decisions based on that, even though it's

18:13

like not very well done. Yeah. And since we've already know how wrong the FBIs data can go, do you think for the 2022 and 23 data, we will still see the same kind of media coverage and people's decision making based on that data? Yeah, definitely.

18:37

I think, since more agencies will report overtime with the estimate should get better, even if you don't actually change their method. Yeah, I think it's still gonna be pretty bad, risky stocks, so much missingness. And he still had an entire states that are reporting, so it's very, not random. And I think the media, especially like the New York Times, Washington Post, like the big national ones, they're gonna be reporting on these numbers and do it pretty accurately.

19:06

Even if they cite like later on in the article like teacher estimates and their concerns about it. I think local news is usually a lot better, because then they can also well, they don't really they don't estimate FBI does not estimate local data. They just do state data. But I feel like local news, understands the circumstances, better of like what's going on in their own town. And they could talk to like the police chief and see like, do these numbers make sense? Or even at the state level, they're probably more likely to check their specific states data

19:41

than just to report on

19:44

occasion. Yeah. And how do you think so like your understanding how, how far are we from from the like, how long it will take for for the whole country's agencies to

20:00

have like a pretty good participant is patient rates and

20:05

reporting rates

20:07

to the neighbors how, like in five years, 10 years, what's your,

20:14

at least five years,

20:16

even on the old system, we never got 100% I don't think we'll ever get 100% We don't really need 100%.

20:25

But I'd say at least like 510 years, probably closer to five restaurants. But estimates aren't really that important anyways, just because crime policy is so local. So like, what happens in one state doesn't really matter to what happens in another state, even like, what happens in one city doesn't really matter. Just because, like, decisions on policing or like mental health services, or even like poverty services, that's going to happen at the city level, I'm not sure it really matters for crime,

20:56

even like state law, is less important than local decisions. Because how you enforce it, it's really at the local level of the police chief says, or the district attorney says we're not enforcing this law, that is not gonna get enforced. Or if they're saying, you know, it's law enforcement, but it's not a priority. That's going to be different behavior, and that's entirely at the local level.

21:20

But if the crime is like, so important at the local level, why did the FBI still put so much effort and the funding into, you know, transitioning into this new new system? And? Yeah, well, so the new system does have local data, so it has agency level data. I don't know why they're spending so much time doing the estimates. I think I would just say like, once we get like 70 or 80%, then we'll start estimating before that our estimates are kind of a waste of time. But I think it's probably just politics, like, politicians expect national data, like the media expects national data and like, if the FBI says, Sorry, we're not gonna give you anything. People were like, FBI sucks. Like, why are we funding them often not doing the thing we need?

22:09

Even though it's not actually going to be helpful at reducing crime? Yeah. And just go back to the previous question, just to clarify. So like you said, it will take at least five years for for the country to reach like, what percent of reporting right? So like, 85? Or 90%? I see.

22:32

Yeah, um, that's pretty much what we have. Do you have any questions? Yeah, actually, as we talked to Josh, and another criminology professor, we have interviewed before, they also believe that, like, by 2025, the country can reach 80% of transition rate to the new system. And by the end of the decade, it is possible for us to,

22:59

like reach to the track where we have seen before when we use the perverse

23:06

data collection system. So do you think it is like a fair prediction, as you said, it may,

23:14

by the end of 2025,

23:17

could happen, they're more optimistic than I am about that.

23:22

But I don't know, I don't, I guess to 2025, it would be

23:28

pretty soon, that's like, two or three years, depending on how you count cuz the data always has lag, about a year before they release it.

23:37

Yeah, think that's actually too early. But

23:42

like, we could get maybe 70% by 2025. But it also depends how you measure it. Because if you're looking at like, the number of agencies that are different than the number of like, officers in those agencies are like the number of civilians who live in those in those jurisdictions.

24:04

And it also matters. If you're looking at like state or national estimates which people want even though it's not that great.

24:12

You really want like a random sample of agencies,

24:16

that will really deal with a lot of issues of non reporting. And that's not really what we see at all. It's usually like some states have a ton of agency, some states have very few.

24:27

I think we're probably going to keep seeing that where it's like a state of being or if like the state says, You're gonna have to report that agencies will start reporting.

24:37

Even like five years from now, we'll probably see something similar.

24:42

I see. All right. Thank you so much. Is there anything you want to add on?

24:49

I assembled link in the chat post.

24:53

Yes.

24:55

Yes, I saw it.

24:59

Yes.

25:00

That's all.

25:02

All right. Thank you so much. Thank you so much. Yeah, I hope you have a good day and good luck with your talk on the conference. Good luck with this project. Thank you. Thank you so much. It's really helpful. Bye bye