Key Insights

Of the 276 municipalities in our data set, there was a wide spectrum of data protection and accessibility standards in place. From a standpoint of protecting personally identifiable information (PII), the municipalities where record requests must be submitted in person with proof of being an involved party in the incident were classified as having the most secure policies. It is likely that many municipalities possess these standards due to a lack of budget, technical sophistication, or both, rather than being motivated by data protection goals. On the other extreme, cities like Dallas, TX, provide instant, no-charge, online access to direct identifiers for victims of crimes to anyone who submits an incident identification value into the query interface. A disclaimer on the Dallas police portal states that the ease of access to this sensitive information is an example of the department's commitment to fostering transparency with the public. From a doxing standpoint, this level of data access and disclosure is an egregious violation of common-sense PII protection standards. Would-be criminals could quickly, easily, and anonymously obtain information that would allow them to re-victimize those who had already been the targets of hate crimes (or any crime, for that matter).

The Good

- Cities within the "A" grade category demonstrate most reasonable balance between data accessibility and protection against re-identification of involved parties in police reports. Police reports are available via an online medium, usually associated with a fee. Incident data that is publicly available for free is without direct identifiers and at least partially anonymized. Some notable cities that possess grades in this category are New York, NY, Los Angeles, CA, and Philadelphia, PA.

The Bad

- Direct identifiers provide nefarious actors with the most basic technological competency the ability to successfully dox exposed parties. Although we prioritize the threat of victim and witness doxing above that of suspects, it is still unacceptable for municipalities to leave suspects vulnerable by exposing their PII to the public. If for no other reason, suspects' sensitive data should also be protected due to the principle that defendants are presumed innocent until proven guilty beyond a reasonable doubt. Municipalities with a grade of "C-" have been shown to expose direct identifiers for suspects. Some notable examples of cities with "C-" grades are Miami, FL, Madison, WI, and Glendale, CA.

The Ugly

- The most egregious violators of common sense data protection policies are those cities who expose direct identifiers of victims, putting them at significant risk for doxing. These are cities who possess a grade of “D+” or below. Municipalities that receive a rating of “D-” and “F” provide full names and home addresses for the victims of reported crimes, often for free and with the ability to access the data anonymously. A short list of notable cities in this category are Dallas, TX, Indianapolis, IN, and Austin, TX.

On Hate Crime Reporting

In order to craft effective mitigation policies to stem the frequency of hate crime, an accurate and comprehensive corpus of hate crime data must exist so that analysts can better understand its prevalence, location, and nature. Unfortunately, our understanding of the issue is only as good as the data provided by local law enforcement agencies to the FBI, who only provide such data voluntarily, since there is no federal mandate. It is no surprise, then, that the FBI shows roughly 6,100 hate crimes per year across the nation, but a Department of Justice survey estimates the number to be closer to 250,000 per year[1]. Many reasons contribute to this massive under-representation of the facts about hate crimes. For one, it is estimated that more than one-half of hate crimes are never formally reported to the police.[2] Further, local police officers are not well-trained to identify and report hate crimes.

It would be unreasonable to recommend that the FBI take on responsibility for governing hate crime reporting down to a local law enforcement agency level, given there were more than 30,000 law enforcement agencies in the US in 2016.[3] A more scalable method of governance may be to work with whichever governing body oversees the accreditation of local police departments in each state, usually affiliated with the Department of State’s office, to facilitate some sort of centralized oversight into hate crime reporting. Perhaps, the Department of Justice could lobby hard for each state governing body to add a new hate crime identification and reporting training to the accreditation criteria. Any police department that does not have a certain percentage of its force trained within 12 months loses accreditation. The Department of Justice could develop the curriculum, training materials, and funding for the state agencies to utilize. Although not a perfect solution, this would ensure that each department at least has some officers that are well-trained in hate crime identification and reporting.

On Data Protection

While understanding the importance of a police department’s desire to foster a sense of trust through transparency by making police reports available, there must be some common-sense data protection policies in place to prevent the exploitation of those listed in official police records. At the very minimum, direct identifiers associated with any involved parties must be redacted from publicly available police records. Ideally, police reports are only made available to

individuals who have been validated as an involved party to the incident. Any data that is released publicly should be aggregated and either void of indirect identifiers or with K-anonymized indirect identifiers.[4]

A policy that might serve as a deterrent to malicious attempts to re-identify individuals from public data would be to criminalize such activity at the federal level, as many European countries have done. The Federal Trade Commission has recently noted the lack of any central government policy regarding the management of re-identification risk in public datasets and the potential of problems to occur as a result.[5] This lack of policy is somewhat surprising, given the detailed recommendations put forward in 2012 by the Computer Security Division of the National Institute on Standards and Technology. Moreover, entities as large as Australia[6] and the United Kingdom[7] have legislation that criminalizes re-identification from public datasets in the absence of a demonstrated public good or other legal, needful purpose (e.g., academic research). It is not the case that the importance of dealing with issues related to re-identification in public datasets, including government-curated datasets, is unknown. The U.S. government has simply failed to act in meaningful way in response to the known threat when it comes to hate crime data. The federal government should consider legislation similar to that of other countries, whereby the structure of publicly available data is regulated and re-identification of crime victims without legitimate need-to-know is criminalized.

Directions for Future Research

Given the importance of possessing comprehensive hate crime reporting data, any methodologies aimed at increasing the frequency of hate crime reporting should be a research focus. Interventions at the local police department level seem to be one of the most direct methods to increase hate crime reporting, short of legislation. Research focused on the effectiveness of different content and delivery media (e.g. seminars, wallet cards, etc.) could have a positive effect on the training void that exists with police officers at the local level.

Endnotes

1. Documenting Hate - ProPublica. (2019, April 12). Retrieved from https://projects.propublica.org/graphics/hatecrimes
2. Confusion, Fear, Cynicism: Why People Don't Report Hate Incidents. (2017, July 31). Retrieved from https://www.propublica.org/article/confusion-fear-cynicism-why-people-dont-report-hate-incidents
3. National Sources of Law Enforcement Employment Data. (2016, October 4). U.S. Department of Justice Office of Justice Programs Bureau of Justice Statistics. Retrieved from https://www.bjs.gov/content/pub/pdf/nsleed.pdf
4. El Emam, Khaled. (2016). A de-identification protocol for open data. In *Privacy Tech*. International Association of Privacy Professionals. Retrieved from https://iapp.org/news/a/a-de-identification-protocol-for-open-data/
5. Cranor, L. (2016). Open police data re-identification risks. Federal Trade Commission. Retrieved from https://www.ftc.gov/news-events/blogs/techftc/2016/04/open-police-data-re-identification-risks
6. Phillips, M., Dove, E. S., & Knoppers, B. M. (2017). Criminal prohibition of wrongful re‐identification: Legal solution or minefield for big data? *Journal of Bioethical Inquiry, 14*(4), 527-539. doi:10.1007/s11673-017-9806-9
7. Data Protection Act 2018, c.12, s.171, s.172. http://www.legislation.gov.uk/ukpga/2018/12/pdfs/ukpga_20180012_en.pdf.