



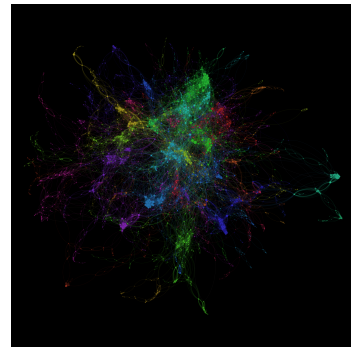
DATAHACK: UNDERSTANDING POLICE MISCONDUCT

THE POLICY LAB & THE JUSTICE COLLABORATORY

Every day a civilian is shot by a police officer in the United States. While rare as compared to other types of violence, the use of force by the police is of potentially greater importance: Force, utilized improperly or routinely, can erode trust between citizens and their government, prompt disengagement with the law, and shake the very foundation of our democracy.

Despite its importance, and the recent attention to police misconduct as a pressing social problem, there is still much we do not know about the phenomenon - including the rates of police misconduct, its distribution, and its root causes. In large part, this is due to a lack of large-scale scientific inquiry, which itself results from a lack of data; neither police departments nor the government are compelled to compile and analyze incidents of police (mis)behavior in any systematic fashion.

The recent series of high-profile shootings of unarmed citizens, however, has drawn the nation's gaze to this issue once again, and citizens, policy makers, and policing experts are mobilizing to uncover what data do exist. One such repository comes from Chicago, IL, and contains records of allegations of police misconduct and investigations of police shootings dating back to 1971. The goal of this DataHack is to leverage this database to answer [the White House's call](#) to develop smarter, more data-driven methods to understand and improve policing in the United States, and to reduce the use of force by officers. Participants will have access to a database - compiled by [The Policy Lab](#) and [The Justice Collaboratory](#) at Yale Law School - containing more than 150,000 records of alleged police misconduct (built on reports made by both citizens and other police officers), as well as related datasets - including departmental and (anonymized) personnel records, maps of the city, and details as to its physical and political infrastructure. Hackers will be charged with uncovering actionable insights into the correlates - and possible drivers - of police



Communities of officers in Chicago

Date: February, 2017.

misconduct, and will be evaluated based on the novelty, robustness, and scalability of their finished product, as well as the degree to which it highlights new conceptual threads as the foundation for future research.

At A Glance: Possible Guiding Questions

Improving our Understanding of Police Misconduct

- Is police misconduct a group or networked phenomenon? If so, what do police misconduct networks *look* like? To what degree do police misconduct networks look and act like social and behavioral networks? For example, do people who work and train together, engage in misconduct together?
- What are the behavioral and network-related correlates of police misconduct? Are there particular characteristics by which we can differentiate officers who do and do not display more serious forms of misconduct?

Improving our Prediction of Police Misconduct

- How accurately can we predict future aggression by police officers? Can we achieve top-shelf accuracy and efficiency in our predictions, but do so using models that can later be interrogated, so as to help policy makers and researchers better understand the underlying behavioral mechanisms at work (i.e. using something other than machine learning)?
- How can we make all of these insights most useful to police departments, in real time? Can we develop visualization tools to aid police departments and citizens in understanding how misconduct concentrates within social networks?

Improving our Study of Police Misconduct

- How can we achieve the objectives listed above, but avoid over-fitting? These data come from one city (Chicago), within which some patterns may be idiosyncratic. Can we build models that are likely to scale to other cities?
- How can we estimate the scale and contents of missing data? These allegations do not represent every incident of police misconduct – rather, they reflect only those that were *officially reported* - a time and energy intensive process, which also requires that citizens trust the police in the first place.

