Lecture 8: Data ethics

Criminology 1200

Prof Maria Cuellar

University of Pennslyvania

Interlude: Correlation vs. Dependence vs. Sampling bias

Correlation: Two variables are correlated if their correlation coefficient is nonzero:

$$\text{Correlation coefficient: } \rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

where $E(X)$ is the expected value of a random variable X, and Var(X) is the variance of X.

Independence: Two variables X and Y are independent when their joint probability distribution is the product of their marginal probability distributions: for all x and y,

$$p_{X,Y}(x,y) = p_X(x)p_Y(y).$$

Two variables are independent when the value of one gives no information about the other.

- Independence implies uncorrelation: If two variables are independent, then they are uncorrelated.
- Uncorrelation does not imply independence: But it does not work the other way around.

Correlation vs. Dependence

Dependence is causal while correlation is associational.

Classic example of dependent data

Suppose there is a trial to test whether a vaccine against covid works.

If an individual is in a group where everyone else is vaccinated (and the vaccine works) then they are less likely to get covid regardless of their vaccination status (i.e., herd immunity.) This makes it harder to test whether the vaccine works for the individual because the outcome is being obscured by the peers.

The status of the peers affect the outcome of the individual.

Sampling bias

The simplest sample for making inferences about a population is a simple random sample.

Simple random sample: A subset of individuals chosen from a larger set (called the population) with the same probability.

- Randomization ensures that this happens.
- Example: taking every 5th person out of the phone book.

There are other types of samples that can be used to make inferences.

But, if a sample is biased (e.g. convenience sample, or one that was collected out of convenience), then we will make an inference that is biased.

Data ethics: Definition

“Data ethics is a new branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values).” - Oxford professors and philosophers Luciano Floridi and Mariarosaria Taddeo

Nice article on Medium: https://medium.com/big-data-at-berkeley/things-you-need-to-know-before-you-become-a-data-scientist-a-beginners-guide-to-data-ethics-8f9aa21af742

Goal of your training

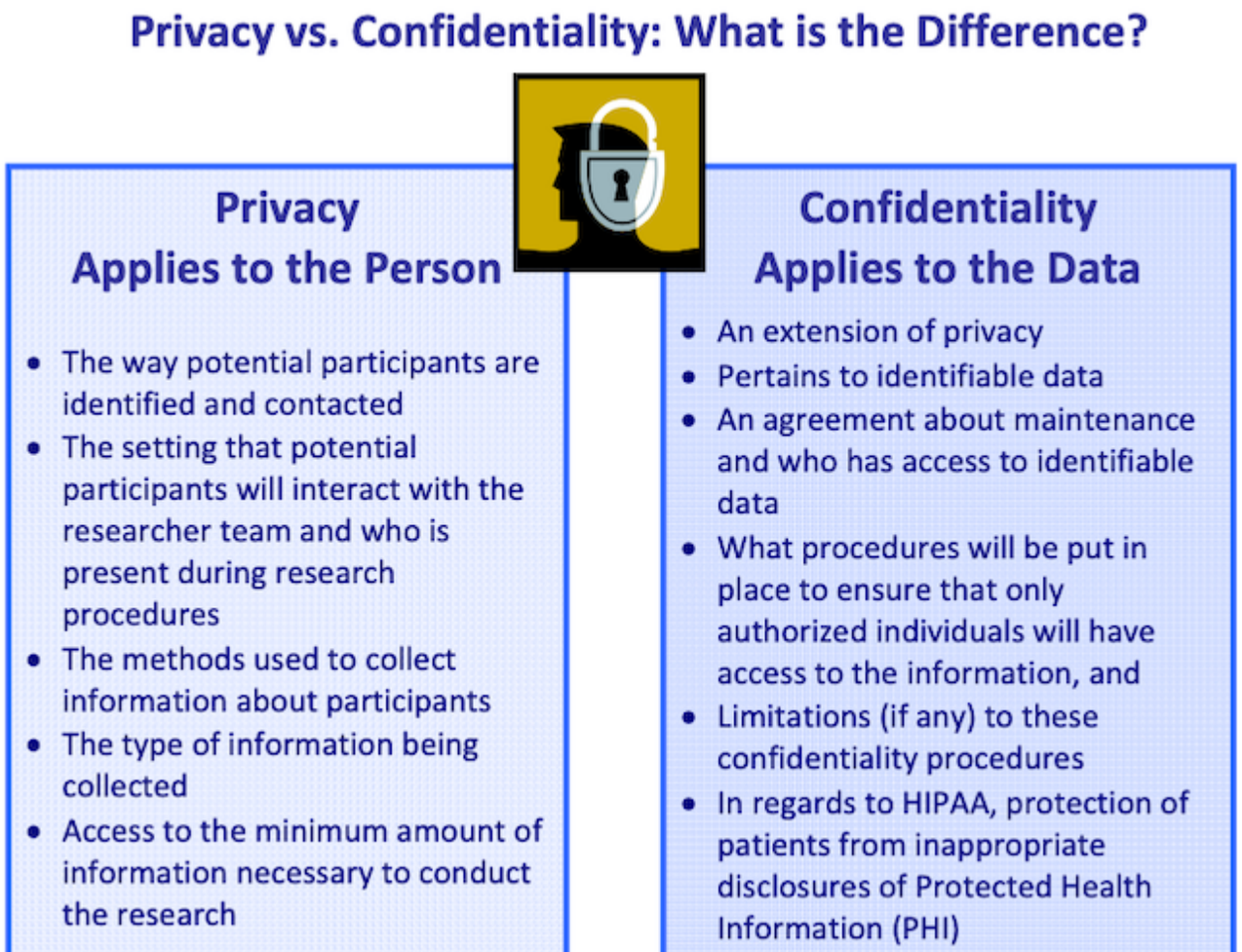
“produce graduates who not only have deep technical expertise, but who also know how to responsibly collect and manage data, and use it to inform decisions and advance innovation to benefit the rapidly evolving world they’re graduating into”.

Statistics really begins before the data are collected

Questions you should consider if you are collecting data, or if you are using someone else's data:

- Population: What is the population of interest?
- Sample selection: Which people should I sample, and which should I not sample? (Cost constraints?)
- Missing values: How should I deal with the missing observations? What do they represent? (e.g. did not want to respond, was not available, their answer was not one of the options, etc.)
- Meta analyzes: If you are using a dataset that was created by compiling several datasets, what is the methodology for each one?
- Consent: Have you informed your respondents that their participation is voluntary? Are the incentives in place coercive?
- Biased options: Even the questions and answers you make available can be biased. Have you checked that these are inclusive and recognize different points of view?

Privacy vs. confidentiality



Data ethics checklist:

- Have we listed how this technology can be attacked or abused? [SECURITY]
- Have we tested our training data to ensure it is fair and representative? [FAIRNESS]
- Have we studied and understood possible sources of bias in our data? [FAIRNESS]
- Does our team reflect diversity of opinions, backgrounds, and kinds of thought? [FAIRNESS]
- What kind of user consent do we need to collect to use the data? [PRIVACY/TRANSPARENCY]
- Do we have a mechanism for gathering consent from users? [TRANSPARENCY]

Data ethics checklist: (continued)

- Have we explained clearly what users are consenting to? [TRANSPARENCY]
- Do we have a mechanism for redress if people are harmed by the results? [TRANSPARENCY]
- Can we shut down this software in production if it is behaving badly?
- Have we tested for fairness with respect to different user groups? [FAIRNESS]
- Have we tested for disparate error rates among different user groups? [FAIRNESS]
- Do we test and monitor for model drift to ensure our software remains fair over time? [FAIRNESS]
- Do we have a plan to protect and secure user data? [SECURITY]

(Loukides, Mason, Patil)

Global trends: Where will data ethics be relevant?

- Chief Privacy Officers can expect ethics to become an explicit part of their role.
- Technology companies will lead the way for U.S. Federal Privacy legislation.
- Sustainable ethics codes will evolve to better address the challenges of a digital world.
- Product excellence and privacy by design will become synonymous.
- Companies will drive to educate policy-makers and regulators about their technologies.

By Barbara Lawler

Given the profound shift in the digital network globally, policymakers must consider:

- What harms are they trying to protect people from?
- What rights do they want to guarantee?
- What problems are they trying to solve?
- What are the privacy outcomes they hope to achieve for their citizens?

(In)famous example in data ethics: Predictive policing

See other slides.

To predict and serve

- In late 2013, Robert McDaniel – a 22-year-old Black man who lives on the South Side of Chicago – **received an unannounced visit** by a Chicago Police Department commander to warn him not to commit any further crimes.
- The visit **took McDaniel by surprise**. He had not committed a crime, did not have a violent criminal record, and had had no recent contact with law enforcement. So why did the police come knocking?
- It turns out that McDaniel was one of approximately 400 people to have been placed on Chicago Police Department's "heat list". These individuals had all been **forecast to be potentially involved in violent crime**, based on an analysis of geographic location and arrest data.
- The heat list is one of **a growing suite of predictive "Big Data" systems** used in police departments across the USA and in Europe to attempt what was previously thought impossible: to stop crime before it occurs.

https://www.youtube.com/watch?v=lG7DGMgfOb8&ab_channel=MovieclipsClassicTrailers

What is predictive policing?

- According to the RAND Corporation, predictive policing is defined as “the application of analytical techniques – particularly quantitative techniques – to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions”.
- Much like how Amazon and Facebook use consumer data to serve up relevant ads or products to consumers, police departments across the United States and Europe increasingly utilise software from technology companies, such as PredPol, Palantir, HunchLabs, and IBM to identify future offenders, highlight trends in criminal activity, and even forecast the locations of future crimes.

Bias in police-recorded data

- Decades of criminological research, dating to at least the nineteenth century, have shown that police databases are not a complete census of all criminal offences, nor do they constitute a representative random sample.
- Empirical evidence suggests that police officers – either implicitly or explicitly – consider race and ethnicity in their determination of which persons to detain and search and which neighbourhoods to patrol.
- Bias in police records can also be attributed to levels of community trust in police, and the desired amount of local policing – both of which can be expected to vary according to geographic location and the demographic make-up of communities.
- Nevertheless, it is clear that police records do not measure crime. They measure some complex interaction between criminality, policing strategy, and community–police relations.

ML algorithms reproduce data with which it's trained

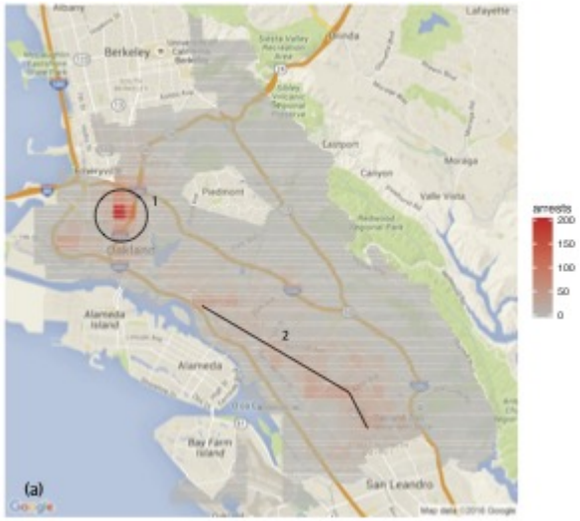
- Machine learning algorithms of the kind predictive policing software relies upon are designed to learn and **reproduce patterns in the data they are given**, regardless of whether the data represents what the model's creators believe or intend.
- Even the best machine learning algorithms trained on police data **will reproduce the patterns and unknown biases in police data**.
- In this sense, predictive policing (see "What is predictive policing?") is aptly named: **it is predicting future policing, not future crime**.
- Because these predictions are likely to over-represent areas that were already known to police, **officers become increasingly likely to patrol these same areas** and observe new criminal acts that confirm their prior beliefs regarding the distributions of criminal activity.

Their approach to finding a "groud truth"

- They combine a demographically representative synthetic population of Oakland, California with survey data from the 2011 National Survey on Drug Use and Health (NSDUH).
- This approach allowed us to obtain high-resolution estimates of illicit drug use from a non-criminal justice, population-based data source which we could then compare with police records.
- In doing so, we find that drug crimes known to police are not a representative sample of all drug crimes.

Drug arrests

Number of drug arrests made by Oakland police department, 2010. (1) West Oakland, (2) International Boulevard.



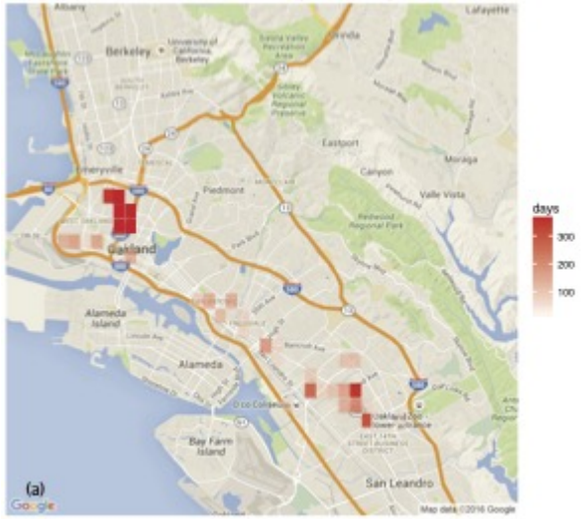
Estimated number of drug users, based on 2011 National Survey on Drug Use and Health.

Evaluate a predictive policing algorithm: PredPol

- Goal: To investigate the effect of police-recorded data on predictive policing models.
- Algorithm: This algorithm was developed by PredPol, one of the largest vendors of predictive policing systems in the USA and one of the few companies to publicly release its algorithm in a peer-reviewed journal.
- Methods: we apply a recently published predictive policing algorithm to the drug crime records in Oakland.

Results

Number of days with targeted policing for drug crimes in areas flagged by PredPol analysis of Oakland police data.



Results

- We find that rather than correcting for the apparent biases in the police data, the model reinforces these biases. The locations that are flagged for targeted policing are those that were, by our estimates, already over-represented in the historical police data.
- Using PredPol in Oakland, black people would be targeted by predictive policing at roughly twice the rate of whites. Individuals classified as a race other than white or black would receive targeted policing at a rate 1.5 times that of whites.

Discussion

- We have demonstrated that predictive policing of drug crimes results in **increasingly disproportionate policing of historically over-policed communities**.
- Although predictive policing is simply reproducing and magnifying the same biases the police have historically held, filtering this decision-making process through sophisticated software that few people understand lends unwarranted legitimacy to biased policing strategies.