LS 123 Data, Prediction, and Law

Meeting 6

**Big Data Policing w/ Sarah Brayne, Mapping Lab 1**

1. ANES Causal Inference Lab Recap
   1. be aware of cases dropped from analysis and coding scheme
      1. Error in recoding education summary var.
      2. Meaning of the values of var's / using interval assumption
   2. means (incl. regression coefficients) are sensitive to outliers
   3. regression results usually fairly robust
   4. note all the assumptions behind regression (in box in lab)
   5. interpretation: model does not account for much variation in DV
      1. be careful of differing scales (rescale 0-1)
      2. using causal models requires theory (here, inductive theories about political psychology)
      3. it is the theory that allows "falsification" of null hypothesis to give you a clue about what is really happening
2. **POSTPONED Guest: Professor Sarah Brayne of University of Texas Austin**
   1. Prof. Brayne is an Assistant Prof. of Sociology at University of Texas Austin. She earned a PhD in Sociology and Social Policy from Princeton and was a postdoctoral researcher at Microsoft Research
   2. Prof. Brayne did extended fieldwork that asked how the LAPD used big data and what difference big data makes for surveillance and police organizations; that research was the basis for what we read and for her book that came out recently from Oxford University Press, *Predict and Surveil: Data, Discretion, and the Future of Policing* (which I just finished and can recommend highly)
   3. she has also studied how contact with the criminal justice system affects people’s other institutional involvements with the financial system, labor markets, education, and medical care
   4. She is is also founder of the Texas Prison Education Initiative
3. **Jon Simon and Malcolm Feeley, from JSP, on “The New Penology”**
   1. the idea that predicting risk and classifying people by the risk they pose would be important predates the era of big data and automated surveillance
   2. our very own Professors Simon and Feeley wrote on it almost 30 years ago
   3. they look at it on a meta level
      1. the discourses of law used to be clinical diagnosis (even in context of rehabilitation) and retributive judgment (which is moral)
      2. those are individualizing frameworks
      3. the shift has been to language of probability and risk
      4. that’s all about central tendencies and aggregations of people
      5. along with shift in language has come shift in purpose, from individual to system, especially efficiently managing systems like society or the carceral system
      6. the language is new but we can ask as we continue in this course (and even when we get to 18th century England)—how new is the penology? was efficient management of risk not on 18th century judges and MP’s agenda?
   4. focus on identifying, classifying, and managing groupings based on dangerousness, both inside and outside prisons
   5. they say the same thing had already happened with tort—from individual responsibility to improving public safety and mitigating risk (this had been going on since 1960 or so)
   6. perhaps risk management would better serve goals of rationality (means tied to ends, or optimizing use of resources) and accountability of officials
   7. identify, classify, separate, and incapacitate based on risk of reoffense
   8. I would add that at some point the optimization of resources got lost in the shuffle, and that punishment no longer was a matter of prison for high risk and probation for low risk offenders; punishments became irrational even as the language of risk remained
   9. risk turns into a way to justify a way manage a permanent offender population; management is what is left if there is no rehabilitation because the state is managing an “underclass” that is permanently on the margin of society
   10. what big data does may just be to make the identifying, classifying, and disposition of offenders more justifiable by invoking science
4. **Sarah Brayne looks at policing in this societal risk management frame**
   1. her research is fieldwork with the LAPD and examining what it does with predictive tools
   2. LAPD, like many other police depts, dropped location-based PredPol in April 2020, but continues to use Palantir data integration even though it has dropped Operation LASER (Los Angeles’ Strategic Extraction & Restoration Program), which was person rather than location based (network analysis)
   3. changes in practice come with big data analytics
      1. discretionary assessments of risk are quantified
      2. data used to predict rather than react or justify
      3. systematic surveillance extends to large populations
      4. people without direct police contact included in surveillance
      5. data generated by policing, adjudication, and punishment joined with data collected for other purposes (benefits, commerce)
   4. surveillance practices based on large amounts of data have their own momentum
      1. organizations may adopt ML and data integration just because other ones do 🡪 institutional isomorphism
      2. incentive to gather as much data as possible to be used later in unforeseen situations (function creep) [you can even call what we do later in the class with the Old Bailey Online to be a type of function creep]
      3. function creep is easy because data collected privately is very rarely protected (there are protections like HIPAA and FERPA for specific kinds of records), and is available for sale on the market
   5. person-based and place-based predictive methods are both implicated
      1. individual point system that was part of LASER, in which you could get 5 points each for violent crime history, gang affiliation, prior handgun arrest, or being on parole or probation, \*but\* you also get a point for each Field Interview card in the system (so, a feedback loop gets generated)
      2. place based system was PredPol (which we will talk about a bit more next time), based on a model for property crime which was like hot-spot policing: if there was a property crime then there was a higher likelihood of a property crime nearby (but applied not just to property crime)
      3. police did not like being second guessed on where the crime happens to be, and did not want themselves surveilled to see whether or not they were in the PredPol boxes
      4. place-based surveillance can be automated, e.g., with automated license plate readers (which allow you to detect crime and then go back in time to do your search), but license plate readers are not deployed randomly either
   6. using data could reduce inequalities in policing
      1. could replace racial stereotyping as basis for suspicion
      2. could reduce hypersurveillance of minority neighborhoods
      3. could improve police accountability and transparency
   7. instead, data-driven surveillance may do the opposite
      1. deepen surveillance of people already marked by the criminal justice system
      2. widening the “dragnet” to include more people (but not uniformly across society)
      3. causing avoidance of essential social institutions (health care)
      4. giving a veneer of scientific objectivity to what really is confirmation bias
   8. this has important implications for how we think about law
      1. privacy—data gathered by a third party could later be used by the state to investigate crime
      2. Fourth Amendment guarantee against unreasonable search and seizure—data is collected without warrant or even suspicion
      3. widespread surveillance data can become evidence after the fact (“pre-crime”)
   9. we should always ask “what is the process that generated this set of data?” and what the implications are for equal protection, privacy, bias, and so on
5. My reflections on Sarah Brayne’s ASR article and book \_Predict & Surveil\_
   1. it seems as though whatever improvements big data has made so far have been at the margins
   2. institutional isomorphism carries the day—police see military or other police organizations using these tools, and so they adopt them
   3. rank and file police do not seem to be all that engaged by what big data can show them