LS 123 Data, Prediction, and Law

Meeting 7

**SFPD Data and Machine Classification, Folium Mapping II**

plan for today

Guest presentation: JSP PhD candidate Brie McLemore on surveillance

a bit more from me on predicting crime

Folium: choropleth maps and brief discussion

time to talk with your group about the Data Investigation project proposal

note: if you have questions on assignments, the answers to which may benefit other people in the class, please post them on Piazza rather than ask through a private medium or on Discord

1. **Guest presentation: Brie McLemore, PhD Candidate, Jurisprudence & Social Policy**
2. **Jon: UCSD grad course machine learning report**
   1. Reading for this class meeting—someone’s homework! For a grad class at UCSD
      1. Class is on machine learning, so the concerns are more CS than they are about social reality, or what really happened, etc
      2. Open competition on Kaggle.com (win prizes! By coming up with algorithm to solve some problem—e.g., identifying “toxic comments” online
   2. soon to be familiar data: SF incident report dataset
      1. Data contain errors (latitude 90! Not as much fun as Latitude 20! But also wrong) what does Lat 90 even mean? Missing data? Is there a pattern?
      2. Decision on what to do with missing data (in this case? Exclude those cases)
      3. They do some analysis of the training set—alternate weeks (and impute that test set are the other weeks)
   3. Goal is simple: classify the incidents into the SFPD’s categories
      1. What do they use as predictors? [day, hour, month, district/grid section]
      2. What data don’t they use? [incident description] why not? [just more verbose] what could you learn from the description?
      3. Might it be fun to predict “resolution” of the incident? Yes if you are a sociolegal scholar, no if you are building a prediction algorithm for crime
      4. Note that they are not working with any theory of crime, even if they wave their hands at what might explain the patterns in the data
   4. What do they actually do to build a classification algorithm?
      1. They tested several!
      2. Logistic regression (any explanations out there?)
         1. use a transformation to approximate a zero or one prediction, since regression assumes a y that is continuous, unbounded, etc
         2. logistic regression (logit) is actual estimating not the probability directly but instead the log of the odds that something is true (like is in a particular category or not), which gives you a sigmoid function that approaches 0 asymptotically on one side and 1 on the other
         3. this can be extended across multiple categories—multinomial logistic regression
      3. Naïve Bayes: assign a case to a category based on the features of that case, assuming that the features themselves are independent (strong assumption)—you can find a clear explanation in the SciKit Learn Documentation
         1. Review: Bayes’ Theorem
         2. here what we are saying is that there is a category (let’s call it *y*) that we would like to assign a data point to, based on a vector of features (*x*1 , …, *x*n)
         3. it is “naïve” Bayes because we are assuming that each feature contributes independently to the probability of that data point (could be a police call, could be a document) being in the class *y*
         4. in the real world, though, the features are probably not independent but correlated
         5. turns out that does not matter so much for making classifications
         6. so restated, the Bayesian way of talking about the probability of being in class *y*, based on a vector of features x, is (using product notation and assuming the independent probability of each feature given the class y)
         7. or, since the vector of features is constant for a given input, just say that the probability of a datapoint being a member of a category c sub i, given a vector of features x, is proportional to the proportion of the dataset that is in that category times the product of each conditional probability of a feature given that category
         8. calculating the individual probabilities for a feature conditional on being in class c sub i depends on the distribution of the features—if they are word counts then the features would follow a multinomial distribution
         9. for the rest I refer you to other courses and the SciKit Learn documentation
         10. but again, you are making a big assumption that the probabilities of a feature, given membership in a category, are independent; for example if you have a document that is in the category press release, you are assuming that the probability of any word appearing in that press release is independent from the probability of any other word appearing
         11. or, if you have an incident in the category larceny, you are making the assumption that the features (time, location, day of the week) are all independent given the category larceny
         12. clearly this does not reflect the real world even if it is still okay at classifying
      4. Random forest: make a bunch of decision trees of a certain number of ply (do we still use that word?) and then find a central tendency of those trees (either a mean or modal value) to come up with an answer
         1. In this case they used different numbers of trees (often 150)
         2. And they restricted the depth of the trees (often to 20 ply but sometimes less)
      5. We already talked about logistic regression; we will get to Naïve Bayes when we talk about text classification, and return to random forest in early March
   5. Is there a key takeaway from this short report ?
      1. The models do not vary all that much in their predictive power, even by trying to add features (like a more precise grid with smaller boxes)
      2. The proportion of accurate predictions is between 21 and 25%
      3. So the data don’t allow for great classification predictions—why not?
      4. The data are skewed: there is just a lot more larceny (from the pie chart it looks as though, if you always predicted that a crime was larceny, you would probably be right about 20% of the time)
      5. Note that potentially better models (e.g. gradient-boosted decision trees, neural networks) are computationally more expensive and harder to implement (or even explicate)
      6. Other classification models, like k nearest neighbors, are stymied by the distribution of data (because predictions tend to lump into the most popular class, larceny)
      7. Perhaps you could classify more accurately by adding more data—weather data, transit use data, etc
   6. For crime, it seems like the “where” may be more predictable than the “what kind” question
   7. But this paper seemed applicable here—what are we trying to predict and are the predictions useful? Besides, it uses the SFPD incident reports!
3. **Mohler et al, Randomized Controlled Trial**
   1. Brantingham the elder is associated with the environmental theory of crime; this is an experiment driven by that theory (and a refutation of randomizing police patrols a la Bernard Harcourt
   2. the experimenters interest in the outcome of the experiment not just theoretical, but also financial (<https://www.predpol.com/about/company/>)
      1. Jeff Brantingham is still a manager of the firm and is a cofounder; George Mohler is the Chief Data Scientist
      2. does that matter?
      3. depends on whether you consider the paper a vindication—the substantive reduction in crime over the status quo is 7.4% of the total of 58 crimes/wk (4.3 crimes/wk/LAPD division)
      4. this could be a good, low cost intervention
      5. seems to be an improvement over the analyst predictions
      6. on the other hand, there might be other interventions and other ways of saving money that are more effective
   3. design of the research—how did they go about demonstrating an effect?
      1. it is a randomized controlled trial (an experiment): random assignment to treatment and control conditions
      2. focus on certain categories of crime:
         1. in Kent—burglary, car theft, theft from vehicle, criminal damage, violence against the person, robbery (56% of total)
         2. in LA—burglary, car theft, theft from vehicle
      3. treatment: the ETAS algorithm selected the 150m by 150m patrol boxes
      4. control: one of the analysts chose the patrol box
      5. assignment: on any given day, either the treatment or control condition was assigned
      6. note that patrol units in Southwest Div. varied considerably in how much time they spent in the prediction boxes (10K minutes to 1000 min over the course of the 6 months of the experiment) and they seemed to spend more time in the boxes as the experiment continued
   4. so how effective is the ETAS model at controlling crime?
4. **Newark Gang Unit from the Frontline documentary and focusing on the “hot spots”—police are the most difficult organization to change**

**I. backgrounder: Newark PD & consent decree**

policing in the U.S. is coordinate, not hierarchical—in a very decentralized system, the constraints are put in place mostly by federal courts and mostly based on the Constitution

this means that policy and oversight of policing is procedurally very formal, relies on adversary system, which means that only some claims actually make it and get accepted 🡪 this makes policing way more uneven in practice than it is in most industrialized countries (e.g., Japan, but even England too)

not only that, but policing is parceled out to agencies at a very local level—this makes data collection and oversight difficult, and therefore sporadic

also makes local resistance more likely to succeed, perhaps

1. **oversight in a federal system**

Constitution, as the Supreme Court has read it, reserves the “general police power” to regulate behavior to the States

so how has Congress managed to intervene in States’ affairs? two ways:

Commerce Clause power (so under the Court’s relatively expansive post- 1936 interpretation, anything that affects interstate commerce)

power of the purse: Congress can subsidize activity and then threaten to take away the subsidy

* 1. **Civil Rights Act of 1964, Omnibus Crime Control Act of 1968**

Congress uses commerce power and power of the purse to prevent discrimination on the basis of race, color, sex, religion, or national origin by recipients of federal funds

* 1. **Violent Crime Control & Law Enforcement Act of 1994 (42 U.S.C. §14141)**

passed by a Democratic Congress (under Clinton) in wake of beating of Rodney King in L.A.

§14141 uses not the power of the purse but the ability of the DoJ to bring civil enforcement actions—ask a court to declare a police agency in violation, or order changes (under the court’s equity power), when the court finds a pattern or practice of conduct that deprives individuals of rights, privileges, and immunities secured by the Constitution and federal law.

1. **civil enforcement vs. power of the purse**

it was the civil enforcement provisions of the 1994 statute that turned out to be the key to federal oversight of police

this was true even under Bush Administration, but the important point to remember with civil enforcement is that it matters very much who the executive branch policymakers are

1. **more vigorous enforcement under VCCA esp. under Obama**
   1. **DoJ launch 67 investigations since 1995**

of those 24 were closed without agreements, either because the evidence was not sufficient or because the problems were resolved informally

8 settled out of court

* 1. **26 result in monitored agreements**

more than half of these concluded during Obama Admin.

more than half of these were consent decrees approved (and monitored under the supervision of ) federal courts

in 16 of the 26, DoJ found pattern of excessive force (the other 10 were other abuses like racial profiling, failure to investigate sexual assaults)

wide variety of jurisdictions were subject to consent decrees: New Orleans PD, Puerto Rico PD, Seattle, Portland, Detroit, Albuquerque, East Haven CT, Warren OH, Maricopa Cty AZ Sherriff (Joe Arpaio), Ferguson MO, Newark NJ…

some already concluded: LAPD, Washington DC PD

but, not entirely clear that the agreements result in lasting changes in problems like excessive force

plus, difficult to measure their impact

Trump Administration effectively suspended DoJ pattern and practice lawsuits, but Biden Administration has restarted the process

1. **Newark PD’s “police state”**

Newark was a real sh\*tstorm of problems—you can see that in the Frontline episode and in the DoJ report on Newark issued in July 2014 (you can get the Justice Dept report online)

catalog of wrongs that the DoJ claimed not just violate the Constitution’s guarantees but also diminish effectiveness of Newark police

* 1. **unconstitutional searches & seizures**

Fourth Amendment is pretty exacting: to search or seize someone (or evidence) you need to have a specific warrant

Supreme Court has allowed exceptions, like the *Terry* exception—if police have an articulable suspicion that someone is armed and is commiting or about to commit a crime, then they can do a limited search for weapon

this is a weaker standard than the probable cause required for arresting someone (“the crime having been committed”) or for getting an arrest warrant

if it is a proper Terry stop, the Supreme Court has allowed drug evidence found during stop

but you still need to have a proper Terry stop—the rule does not allow police to stop anyone they want just because the person looks out of place, looks like they don’t want to encounter police, etc.

BUT in 75% of the Newark PDs ***own*** Field Inquiry Reports (which were supposed to document each stop) on pedestrian stops failed to articulate a sufficient legal basis for the stop, and that is the proportion of reports that did record a reason—in 16% of stops there was no recorded justification at all

police often stopped people merely for being in high crime areas, which was actually NPD policy

* 1. **disproportionate stops**

the NPD did not do a good job of monitoring whether or not its practices were having a disparate impact on Black and white residents

DoJ investigators had to go to the third party vendor of data services to get the data

e.g. did not collect data for “Quality of Life” citations

of 1500 stops whose only recorded basis was “suspicious person” 81% were Black (in a population that is 54% Black)

if you are Black in Newark you were on average 2.5 times more likely to be stopped than if you are white, 2.7 times more likely to be searched, and 3.1 times more likely to be frisked

evidence recovery during a frisk? 14% for white subjects, 13% for Black

so there are many additional stops but not more evidence of crime for Black people in Newark

* 1. **retaliatory actions & excessive force**

“contempt of cop”—evidence that Newark Police are arresting people for legal behavior, like challenging or objecting to police action

those sorts of expression are protected by First Amendment

arrests were for offenses that did not have a basis in fact—things like resisting arrest, obstructing administration of law, disorderly conduct

ineffective system for even reporting police use of force; in 30% of incident reports that recorded use of force, the officer did not complete the required Force Report

failure of supervisors to review use of force by officers

true even for use of deadly force: all 29 officer involved shooting cases from May 2010 to Jan 2012 had incomplete investigations

DoJ even found a pattern of theft from civilians (esp. from arrestees, who don’t get their property back when they are released)

discouraging civilian complaints by reading Miranda warning to complainants (who are often those who have been arrested) and even witnesses…

1. **consent decree mandates**
   1. **community engagement**

creation of civilian oversight board

implement problem-oriented policing

annual community survey

* 1. **adherence to constitutional requirements**

revise policies and training to ensure stops, searches, & arrests comport with Constitution

better supervision of stops, searches, arrests

respect 1st Am. rights of onlookers and bystanders

* 1. **better reporting and management**

improve training for use of force, including de-escalation

better reporting and investigation of use of force

equip patrol cars with cameras, require officers to wear body cams to record enforcement activity

improve data systems

and on and on

**Ras Baraka—police are like other institutions in society, but with guns (**other institutions like health care have same institutionalized racism, but it does matter that police have guns, and it does matter that they have autonomy and political weight, so they are like other institutions but maybe not exactly the same)

1. **Folium mapping**
   1. choro means location, plethos means a great many, and so a choropleth map represents a quantity by location, e.g., an unemployment rate by state, a Covid infection rate by county, etc.
   2. being able to make visualizations like **choropleth** (I always make it sound like chlorophyll) maps is a) useful, b) marketable, and c) helpful for understanding empirical phenomena like unemployment, public health, voter turnout, you name it