Chicago hosts Strategic Decision and Support Centers (SDSCs) in all but one of their 22 districts. These are the CPD’s real-time crime centers, launched in 2017, that were created to employ data-driven management strategies to address gun violence and crime in its communities.

[Reading Notes](https://docs.google.com/document/d/1Ljc87_A27bsbQY7QyuLzis4Q7TAKRv98TGYKJ76fJ6w/edit?usp=sharing)

* Comparative data analysis on incarceration rates/trends and/or criminal justice policies
* <https://www.vera.org/downloads/publications/a-technical-guide-to-jail-data-analysis-report.pdf>
* <https://onlinelibrary.wiley.com/doi/full/10.1111/lapo.12113>
* <https://www.researchgate.net/profile/Justin-Pickett/publication/325575112_Public_Opinion_and_Criminal_Justice_Policy_Theory_and_Research_Annual_Review_of_Criminology/links/5c3e106b458515a4c7280ed4/Public-Opinion-and-Criminal-Justice-Policy-Theory-and-Research-Annual-Review-of-Criminology.pdf>

Questions:

* Should i typically schedule a separate time outside of office hours to chat about thesis?
* What are some timelines?
* Is there anything you like to do with other students you’ve had for thesis in the past?
* What are the expectations you have of me in terms of timeline or anything else.
* I hope to focus more on the data science and algorithms part of the subject
* Recommend that we set up a weekly meeting time (5-10)
* 2:30 - Thursday meeting with Huber starting next week
* First two weeks - do the scholar.google.com searches
  + Look at papers, look at papers they cite or are cited by them
  + Get an idea of what is out there in the literature
* After, build an outline of what i hope to accomplish
  + Figure out what data i need to make this work
* Next couple of weeks
  + Can this data be easily accessed
  + Library and librarians are a great resource
* After finding databases - can start research
  + Change the topic
* Start research
  + Keep an r markdown (overleaf)
  + Have an outline plan
  + Start putting in code
  + Start visualize, modeling, transforming
* Spring
  + Fill in sections with more detail
  + More of literature to fill in cracks in knowledge
  + Write up what you have found, write up an analysis
  + Write conclusion, introduction
  + Huber can look things over and give final feedback
  + At least 2 weeks before final deadline
* Expository - what is the state of this situations
* Finding a dataset - that you can apply techniques too and analyze results, harder
* Go deep on one of these algorithms, use simulated data, see how easy it is to get daatsets
  + Data on how widely used a particular software is use
  + Focus on a city or county
* [Primer on Criminal Justice Risk Assessments](https://crim.sas.upenn.edu/sites/default/files/WP2016-03_Berk_RiskPrimer_07.10.2016%281%29.pdf)
  + “For example, risk assessments can help inform decisions at arraignments about whether to detain an offender before his or her next court date”
* Data-Driven Justice Initiative
  + Some counties using data-based risk assessment tool to identify low risk people in jail and find ways to release them safely
  + “Jail population has gone down 40 percent, more low risk individuals have been release from jail, there has been no increase in reported crime”
  + What is it?
    - “A bipartisan coalition of city, county, and state governments who have committed to using data-driven strategies to divert low level offenders with mental illness out of the criminal system and to change approaches to pretrial incarceration so that low risk offenders no longer stay in jail simply because they cannot afford a bond” ([white house archives](https://obamawhitehouse.archives.gov/datadrivenjustice))
    - Goals:

1. Combine data across criminal justice and health systems to identify individuals with highest number of contacts with police, ambulance, emergency departments, and other services
2. Equip law enforcement and first responders to enable more rapid deployment of tools, approaches and other innovations
3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Ideas

* Evaluating data based risk assessment tools in incarceration and criminal justice/policing
  + Evaluate the algorithms themselves, the datasets they are based off of, effectiveness, the difference in how these tools are carried out in counties throughout the country (if other countries are using similar tools)
  + Evaluate if the results could be due to policies and regulations that were implemented around the same time
  + Breaking into the “black box” of policing and incarceration algorithms
* Predictive Policing - method of law enforcement that uses data analysis and machine learning algorithms to identify areas and times where crimes are most likely to occur/ in order to predict and help prevent future crimes
  + Analyze how the ethical implications, how far current countries use these algorithms
  + Suggest limits and public policy considerations for law enforcement when utilizing these tools
  + Analyze these algorithms with the consideration of the bias inherent within the original datasets, the history of policing and how these algorithms may exacerbate systemic discrimination
  + From a more mathematical/ data science perspective
* Evaluating Data based tools and algorithms in U.S. Criminal Justice Systems
  + Evaluating both incarceration and policing algorithms, since they work hand in hand
* Check the citations in the work
* Look at what papers are out there and build a bibliography

**"Evaluating Predictive Policing Tools and Risk Assessment Models: A Multi-Algorithm Approach to Crime Prevention in Major U.S. Cities"**

* research would assess the effectiveness of different predictive policing tools (e.g., PredPol) and risk assessment models (e.g., COMPAS) across several major U.S. cities. The focus would be on comparing their accuracy, impact on crime rates, and potential biases.
* Alternative focus could analyze the role of various algorithmic models in how they have shaped criminal justice policy. The study would compare how different risk assessment tools and predictive policing algorithms influence policy decisions and their implications for justice.
  + This study would compare the effectiveness of crime prediction algorithms (like PredPol) and risk assessment tools (like COMPAS) in shaping community policing strategies. The research would examine how each type of algorithm influences police practices and community interactions..

**"The Interplay Between Risk Assessment Tools and Predictive Policing Algorithms: Impact on Arrest Rates and Sentencing in the U.S. Criminal Justice System"**

* explore how the use of risk assessment tools and predictive policing algorithms interact and influence arrest rates and sentencing outcomes. The research would analyze data from jurisdictions using both types of algorithms to assess their combined effect on criminal justice processes.
* Much harder to find data on

**"Algorithmic Fairness in Criminal Justice: A Comparative Study of COMPAS, Risk Terrain Modeling, and Algorithmic Bias Detection Tools"**

* **Description**: This study would examine the fairness of different algorithms used in criminal justice, including COMPAS (for recidivism prediction), Risk Terrain Modeling (for crime prediction), and various algorithmic bias detection tools. The focus would be on how each algorithm addresses or perpetuates fairness issues and racial biases
* Current/System they are phasing out is the summary reporting statistic
* New phasing in NIBRS (National Incident Based Reporting System) - captures much more detailed , location time of day, whether incident was cleared
  + All 50 states complied as of may 2024
  + Los angeles begun March 7. 2024
* Bring back to aggregate - bring back to

Datasets

City of Los Angeles

* Crime data from 2020 to present
* Adult probation
* Automatic record relief
* Crimes and clearances
* Juvenile court and probation statistical system
* [data portal](https://openjustice.doj.ca.gov/data)

Chicago

* Crimes 2001 to present - City of Chicago
* <https://data.cityofchicago.org/>
* Strategic Subject List - Historical
  + Program ended in 2019, shows listing of arrest data from 2012 to 2016, used by CPD
  + S strategic subject algorithm, reflect an individuals probability of being involved in a shooting incident as either a victim or offender

New York

Houston

* Incarceration data/information is available through harris county data
  + Incarceration is run by county
* Crime statistics are from the houston police department
  + Crime and policing done by city

Phoenix

* Policing by city
* Incarceration run by Maricopa County
  + Data collected by Arizona Department of Corrections, Rehabilitation, and Reentry
  + Includes phoenix data as part of maricopa county

Philadelphia

* City of Philadelphia has data on crime and prison population

Portland

* Very extensive policing statistics
* No incarceration data

Atlanta

* Open police data

National Data

* Bureau of Justice Statistics - only goes back to 2022
  + Correctional Populations - summarizes data on populations supervised by probation, in jails
  + Aggregate of the entire country
  + Recidivism of state prisoners - latest 2012
* Was trying to find out if there was a standardized method of data collection centralized in the government, they do, however they are reported directly from states and aggregated to the entire state
* Incarceration
  + List of cities with Incarceration Data - Seattle, Philly, NY, Los Angeles

Questions

* Houston - Jurisdiction of Harris County versus Houston City

To Do

* Look into each individual city and determine what their involvement/use of different algorithms has been
  + Build a kind of profile for each city, detailing criminal justice/incarceration histories, policies, laws, anything in particular that stands out
* Look through the Datasets and start thinking of how to analyze them

- many cities have implemented algorithms and then walked them back already due to public backlash or it didn't do anything to help crime rate and money

- many cities also develop their own algorithms, not using larger well known ones by private companies

- cities are at many stages of implementing algorithms into criminal justice (some have just signed, some have already used for a couple of years and gotten rid of them (not completely))

- too broad, hard to find a commonality to really dive into, or should i just start looking at datasets and seeing what exactly i can pinpoint

- or would i like that to be part of the thesis, illustrating that different cities are at different stages of their implementation of algorithms with policing

- drop houston??????

-can apply to houston

- narrow down dataset - look at one variable of crimes, violent crimes, burglaries, look at one or two statistics

- los angeles - look into those specific studies

-

I have all this data and information and i dont know what to do with it

* General trends
* Specific direction
* A lot of data, its very messy, not cohesive, every city captures something different does things on their own way
* End of semester - know what datasets you are going to be using - have something that you know is an interesting effect
  + Have a plan for what your thesis is going to be
  + Include the code in the paper - in an appendix
  + Put in a lot of comments - paragraph of text, code
  + Data analysis -
* What concrete things am i shooting to complete by the end of the semester

Starting again - 10/17

# Stop and Risk: Policing, Data, and the Digital Age of Discrimination

* Most empirical studies are field trials - trails run on events that have already happened
* What to do in studies where it is extremely difficult to control for variables

Reframing Project

* Start with Illinois - so much data (too much data?)
* Start with Parole and Bail Trends, controlling for any changes in the counts in the different classifications of crime (violent, etc.)
* Recidivism is currently defined in
* Can you make thesis exposition - what data is out there, what data is missing
  + Some municipalities have decided

10/31/2024

* Chose to look at illinois because of the expansiveness of their data
* Look at trends in incarceration and policing (types of crimes)
  + RIPA in chicago - required data collection on complaints and traffic stops
* Qualifying - pandemic trends
  + Outline - introduction
    - Qualitative
    - Methodology
    - Datasets
    - Analysis
    - Conclusion
  + Data exploration - don't need a hypothesis -

The tulsa pre-k study has a crazy way of getting demographic crime rates in Tulsa (pg 17-18)

This week

* Density distributions
* Age group of the top 50 individuals on the SSL list
* Researched into the background of SSL, how they were actually implementing it,
  + 5 different interactions - no direction on what do actually do with it

Next Steps

* I want to look more into district by district effects
  + If there were a change in COPA cases/complaints about officers
    - Could be used as an indicator of police behavior at the time, increased surveillance, increased police brutality, etc.
  + Increased in arrests for each type of crime by area
  + If there was a change in crime around each new iteration of the SSL
* If areas with higher levels of crime correspond to areas where the top50 individuals in the area on the SSL had higher median/mean score, “meaning they had a higher propensity to violence”
  + If true, this would support the theory behind predictive policing
* End of semester goals
  + Intro/history/background written up
  + Dataset section written up
  + Initial analysis written up

<https://idot.illinois.gov/transportation-system/local-transportation-partners/law-enforcement/reporting/illinois-traffic-and-pedestrian-stop-study/studies.html>

<https://data.cityofchicago.org/Public-Safety/COPA-Cases-By-Involved-Officer/ufxy-tgry/about_data>

<https://idot.illinois.gov/transportation-system/local-transportation-partners/law-enforcement/reporting/illinois-traffic-and-pedestrian-stop-study.html>

<https://www.acgusa.org/wp-content/uploads/2020/03/2020_Predpol_Peteranderl_Kellen.pdf>

<https://igchicago.org/wp-content/uploads/2020/01/OIG-Advisory-Concerning-CPDs-Predictive-Risk-Models-.pdf>

<https://www.chicagopolice.org/violence-reduction-strategy-vrs/>

TOC Split

Version 1

Why the heat list?

* Different easy police can predict criminal activity - one is by identifying locations/times, another is by identifying the actual individuals
  + Individuals are already being identified as having potential to engage in criminal activity through non algorithmic means (parole, etc.)
* This thesis focuses on the potential harms and effects of creating/utilizing algorithms that focus on identifying potentially “dangerous” individuals
* Algorithmic use in policing is not heavily researched, especially individual focused algorithms
  + PredPol is the most common geolocating algorithm that is currently being used in precincts across the nation
  + As of currently, there are no known individual level identifying algorithms being used in police precincts, the most widely publicized was Chicago’s Heat List.
  + The heat list garnered much national controversy throughout its introduction and use. … Throughout its use, various policy organizations and the CPD itself conducted operational analyses.
  + It was quietly shut down in \_\_\_\_\_\_\_\_. And since then, there have been no major research efforts to analyze how and whether the Heat list impacted policing and crime in chicago. The thesis attempts to fill the gap, analyzing the algorithm from a qualitative and quantitative standpoint post mortem.

1. Introduction
   1. Policing algorithms - 2/27
   2. Background on Heat list - 2/27
2. Qualitative analysis - Inspector General’s Report, on the ground reports of how the heat list was used - 3/6
3. Algorithm Analyses
   1. Recreate the algorithm?
4. Dataset Analyses - 3/27
   1. Crime Trends/Arrest Trends/Policing Trends
5. Conclusion - 4/10
   1. Touch on other policing algorithms

First full draft - including formatting - 4/17

Question: National Reports do not align with Attorney General Reports and RAND corporation reports. Should i go into these inconsistencies and highlight them?

* How should i be formatting my sources as im actually writing,
* How do I in text citation the national institute of justice, or the CPD
  + Full name acronym, ymd
* Using “I” or “We” in thesis - pronoun usage in thesis
  + Better to write in third person, the next step

3/5/2025

Come up with General Crime Statistics

* What is the state of crime in Chicago before the use of the Heat list in comparison to federal averages? To their own average?
* What about after the quiet dissolution of the heat list?
* What are district level trends in crime? Are there neighborhoods with more violent crimes, or crimes in general?
* Are there seasonal trends? (monthly, quarterly)

District Level

* Pull out the top 20 Heat list individuals for each district - which district has the highest average? Does this line up with the district level trends in crime?