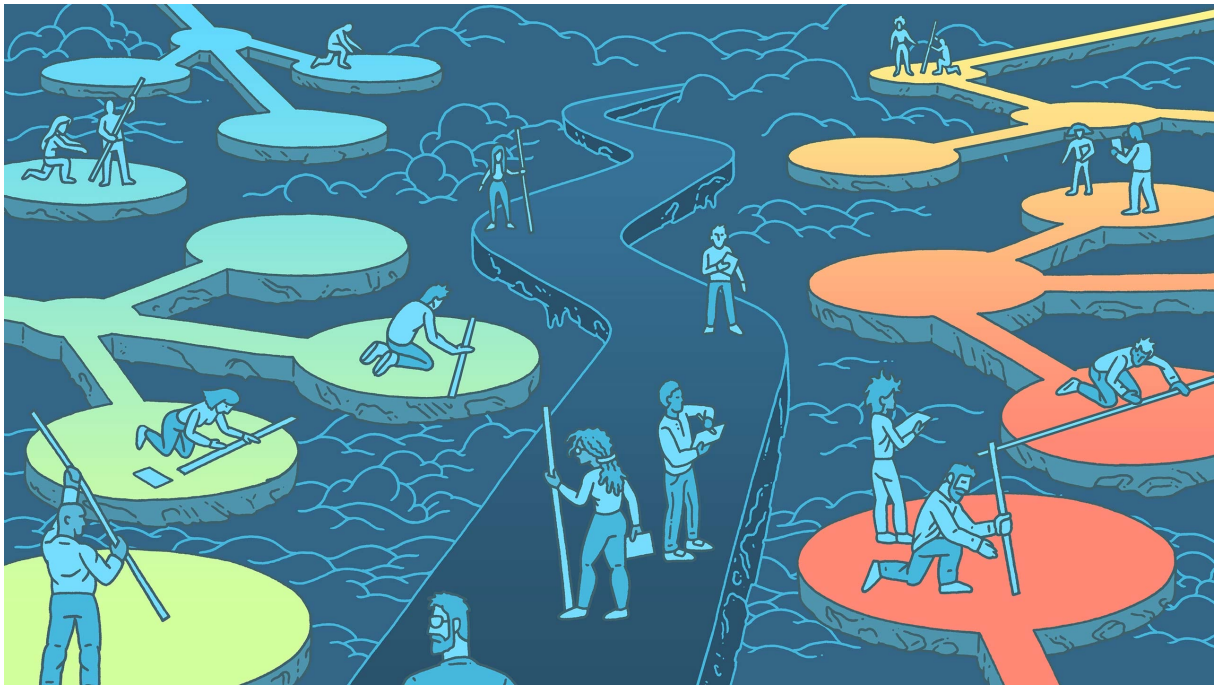
Priority Queues: The Ethics of Classification, Sorting, and Ranking Algorithms

KATIE CREEL

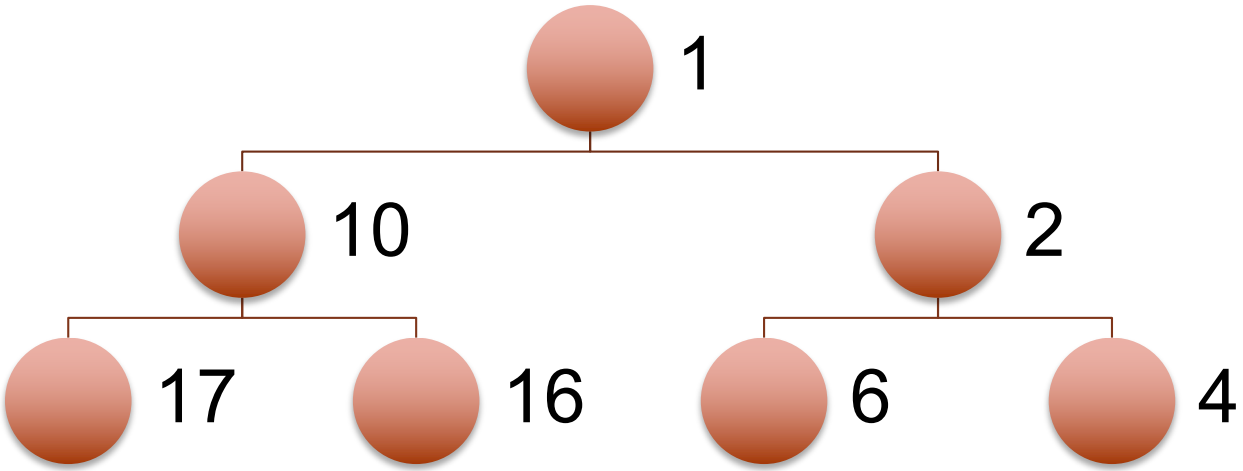
February 16th, 2022



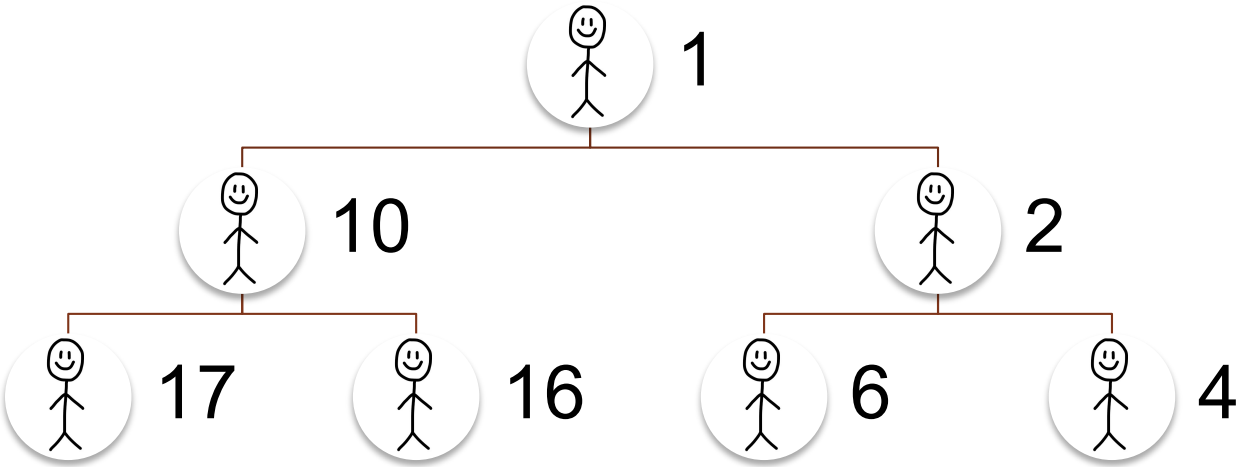
Sorting, Ranking, and Prioritizing Data about Humans

… And Its Problems

Priority Queue (implemented as heap)



Priority Queue ...
of People's Housing Needs



Los Angeles County Coordinated Entry System (CES)

An electronic registry of unhoused persons who are applying or have applied to housing support programs offered by Los Angeles County.

Reference: https://www.lahsa.org/ces/

How does it work?

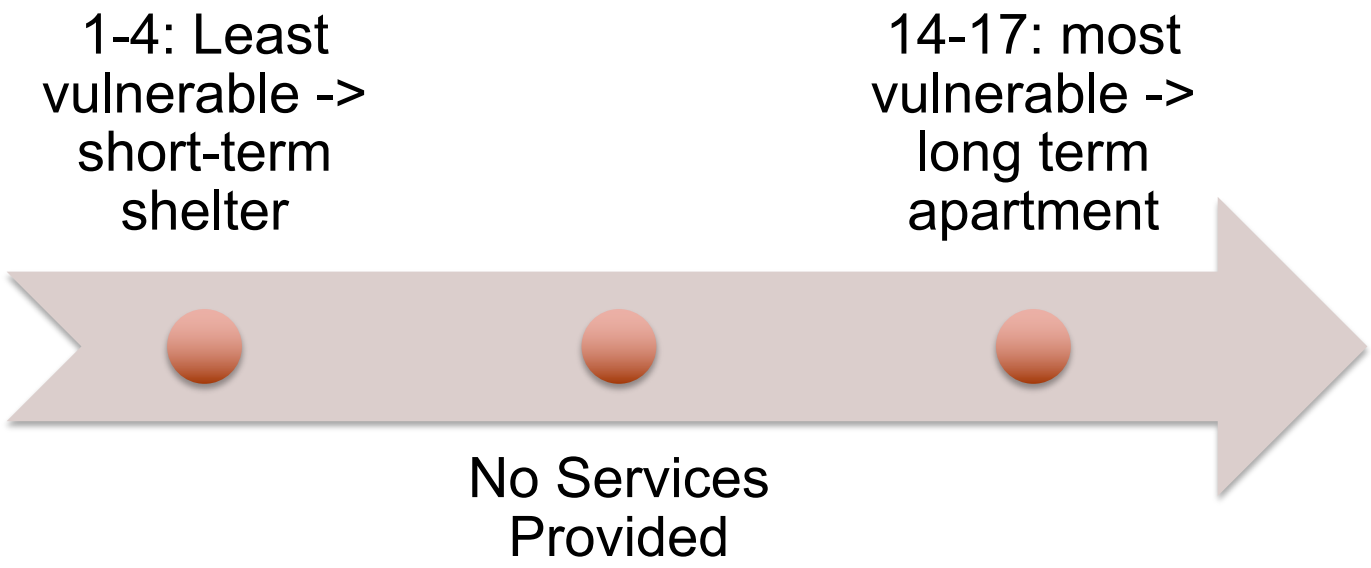
Unhoused person provides (very) personal information including name, DOB, immigration status, current & past mental health, sexual activity, substance use.

Algorithm uses personal data to assign a number from 1-17, least vulnerable to most vulnerable.

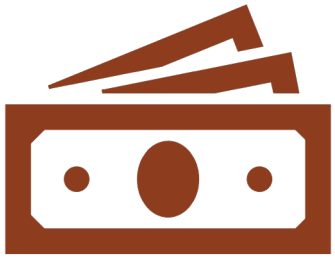
Risk score is used to prioritize and assign housing and housing related services.

Reference: Eubanks, Automating Inequality

[Simplified] CES Binning System



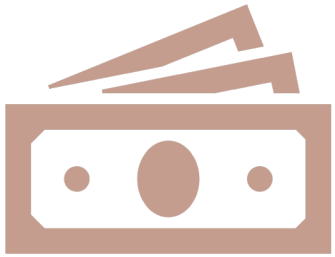
1. Creating Algorithmic Systems Require Resources



Total cost of the CES to LA County: $11 million

Cost per person who has been housed with it: $1,140 per person

An algorithmic prioritization system requires (expensive) engineers to build it, social workers to collect data for it, and tech people to maintain it.

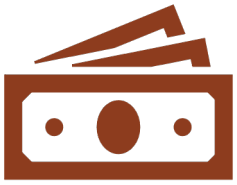


Total cost of the CES to LA County: $11 million

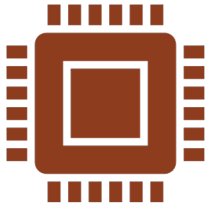
Cost per person who has been housed with it: $1,140 per person

Better Prioritization vs More Services

The CES did improve matching between people and services, but didn't increase the number of people housed.



Would the $11 million have been better spent on giving each person $1,140 to put towards a security deposit for an apartment?



Should this system have been built? How would you decide?

Virginia Eubank's Two Questions for Automated Decision-Systems

"Does the tool [or algorithm] increase the self-determination and agency of the decision subjects?"

"Would the tool be tolerated if it was targeted at non-poor [housed, etc] people?" (Eubank 2018)

If the answer is no, Eubank argues that we should reconsider the design or consider not building the system at all.

What is Autonomy?

"Individual autonomy is ... the capacity to be one's own person, to live one's life according to reasons and motives that are taken as one's own and not the product of manipulative or distorting external forces, to be in this way independent ...

to govern oneself, to be directed by considerations, desires, conditions, and characteristics that are not simply imposed externally upon one, but are part of what can somehow be considered one's authentic self" (Christman 2020).

Discussion Question: How might the CES prioritization algorithm diminish autonomy, self-determination, and agency?

2. Algorithms need Data (but storing data incurs risks)

The collection of data itself is demanding for the data subjects.

- In addition to name, DOB, and demographic information, they are asked for very personal information, like immigration status, history of mental health and substance use.

Generating and storing a large pool of data about a vulnerable population itself incurs risks.

- What if the data leaks or is hacked?
- What if other agencies like ICE attempt to access collected data, such as immigration status + where the person is during the day?

3. What happens to people in the middle?

Survey data “expires” every six months (although the applicants and their data are not deleted from the system!)

This means the people in the middle have to re-take the survey every six months.

What happens to people who are asked to fill out an invasive survey multiple times over a period of years without ever being offered any services?

How might people who are not allocated housing react to this kind of algorithmic sorting?

1. “Game the system” by figuring out how factors that make up the ranking algorithm are weighted and changing them.

Compare this with “Goodheart’s Law” as summarized by Marilyn Strathern:

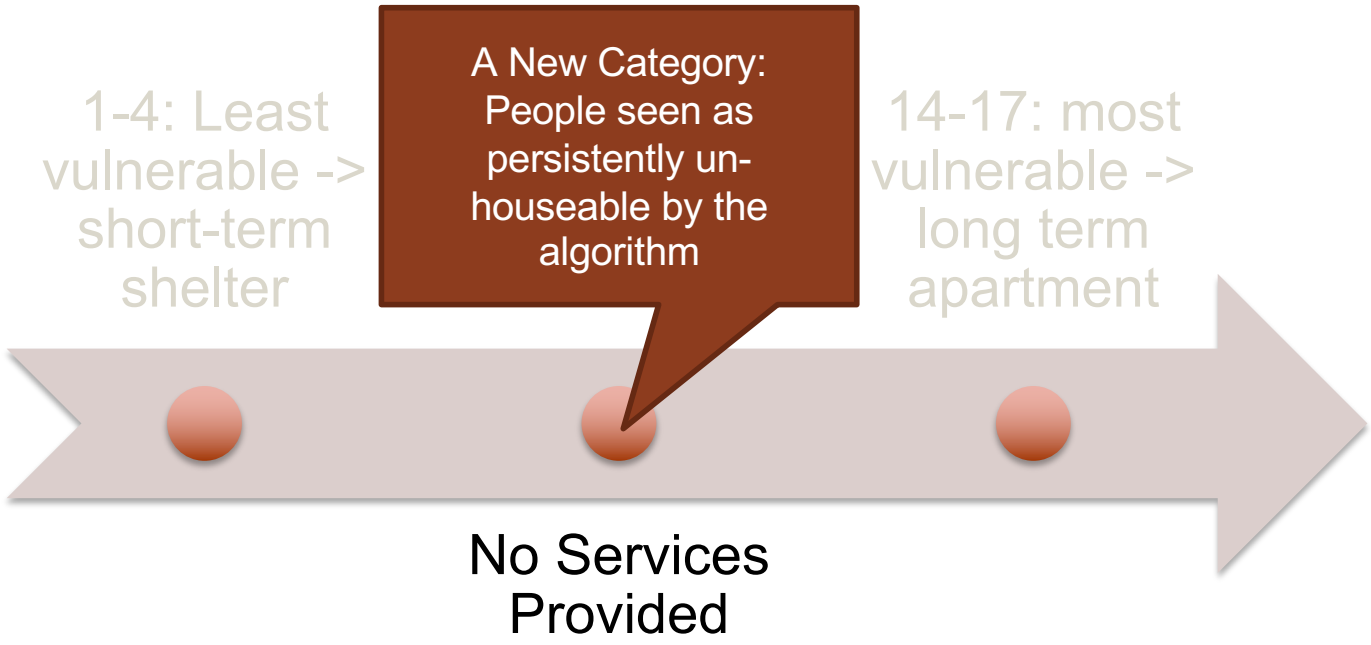
"When a measure becomes a target, it ceases to be a good measure.”

How might people who are not allocated housing react to this kind of algorithmic sorting?

1. “Game the system” by figuring out how factors that make up the ranking algorithm are weighted and changing them.

2. Become cynical and disengaged, refusing to fill out the survey or interact with social workers or other people who claim to be “here to help”

Classification changes people

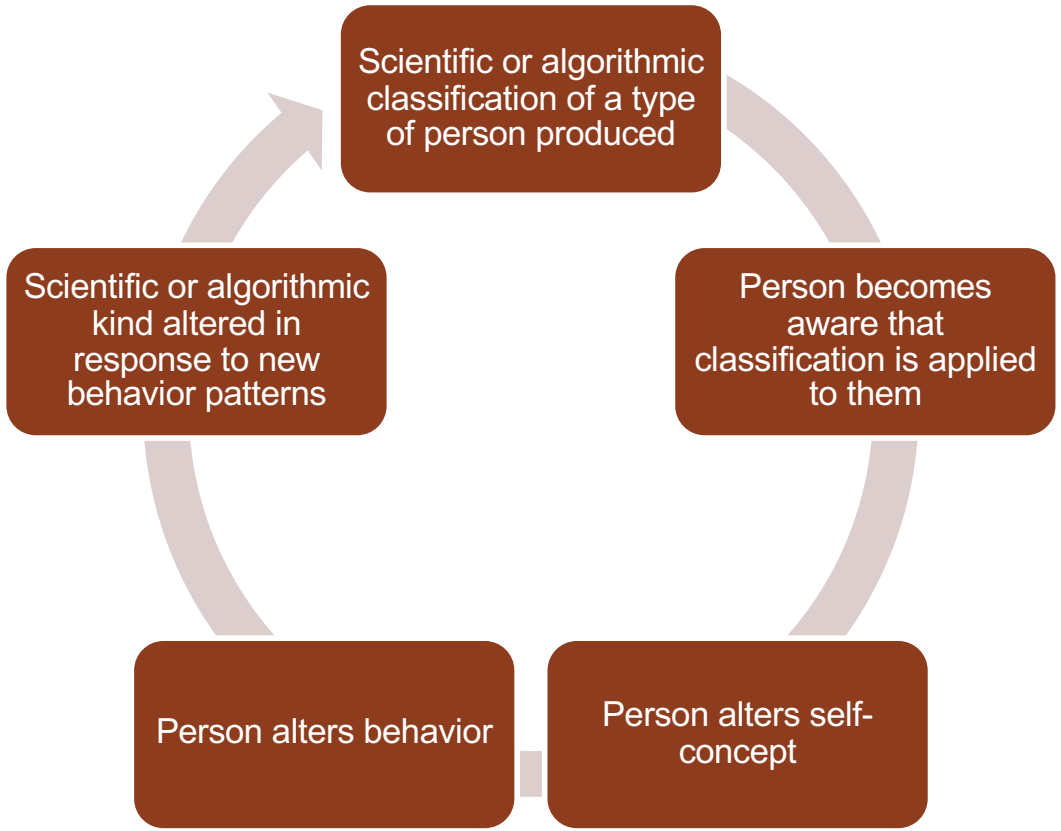


Looping Kinds

The creation of a new category of person, whether by social scientists or by an algorithm, can influence the behavior and self-conception of people sorted into that category.

Source: Hacking, Making Up People

Looping Kinds

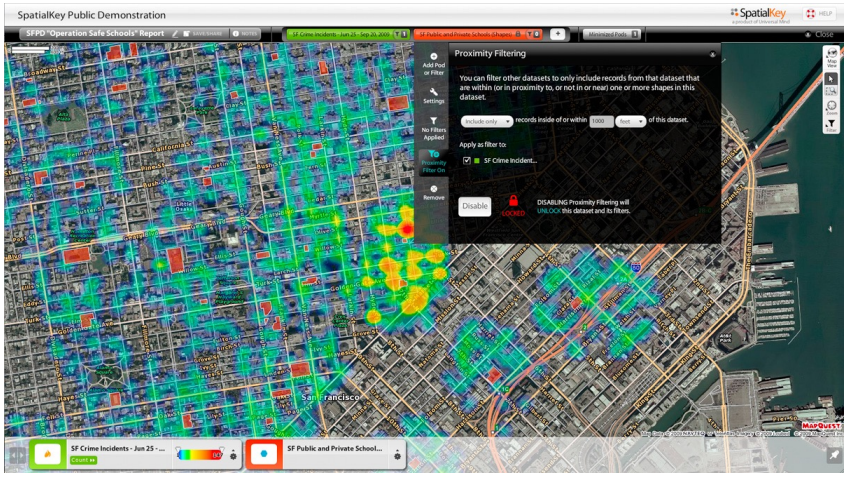


Sorting Individuals

- What are examples of new categories created by algorithmic sorting?
- Have you experienced algorithmic sorting on Tik Tok or another platform, or as the subject of automated decision-making?
- Have you seen a content creator or other person act differently once they become aware of how they have been algorithmically classified?

Sorting Neighborhoods: Predictive Policing

Predictive policing purports to use data about past (observed) crimes to predict locations of future crimes



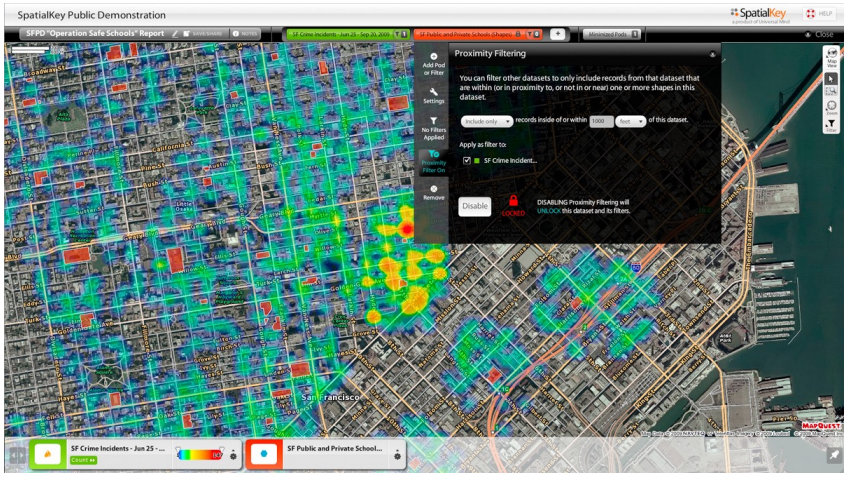
Predictive policing "hotspot" map

Source: https://harvardcrcl.org/minority-report-why-we-should-question-predictive-policing/

Sorting Neighborhoods: Predictive Policing

There are many problems with predictive policing:

- Its efficacy at reducing crime has been disputed
- It is certainly biased: drug use is uniform across racial groups, but “black neighborhoods would be targeted [by the algorithm] at twice the rate of white neighborhoods for drug crimes”
- The algorithms used are proprietary and not publicly auditable.

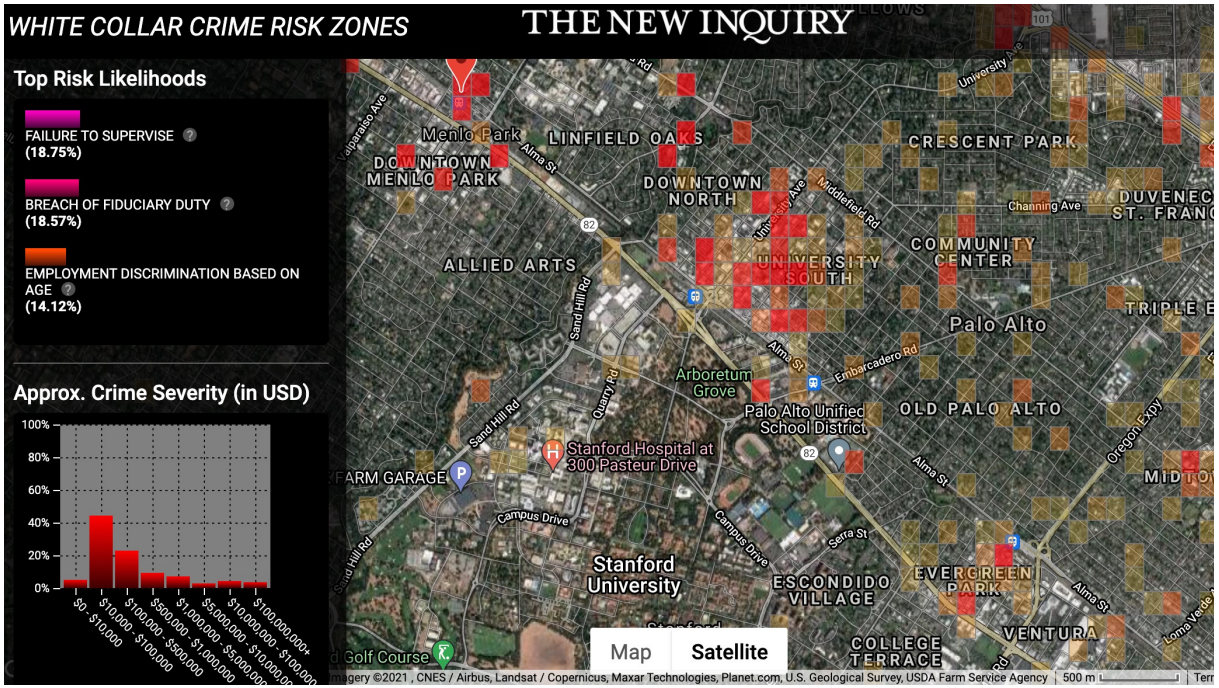


Predictive policing “hotspot” map

Source: https://harvardcrcl.org/minority-report-why-we-should-question-predictive-policing/

For more information: Predict & Surveil by Sarah Brayne

Stanford University



White Collar Crime Risk Zones Map of Palo Alto

Predictive policing typically targets “street crime” and sends police officers to low-income neighborhoods, but the same methodology could be used to predict white collar and financial crimes, which often occur in financial districts.

Source: https://whitecollar.thenewinquiry.com/

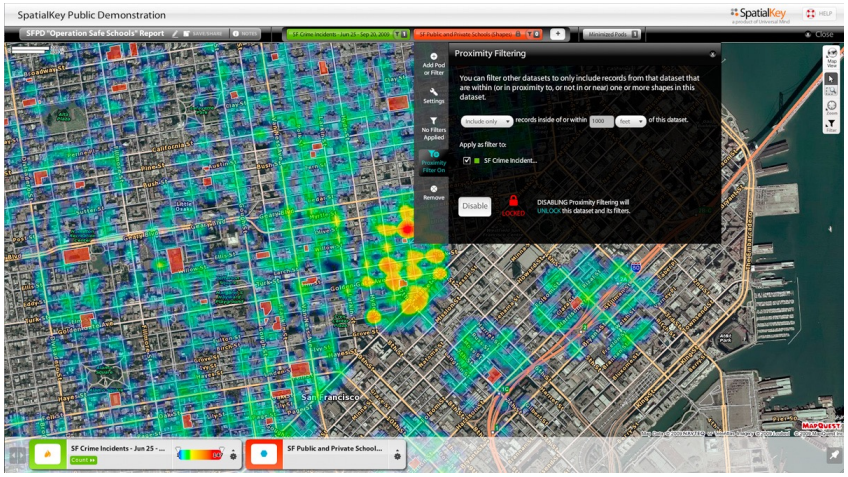
Sorting Neighborhoods: Predictive Policing

For our purposes however, let's focus on the basic sorting mechanism that underlies predictive policing.

Imagine a city in which crime was perfectly evenly distributed but some neighborhoods were arbitrarily labeled high crime.

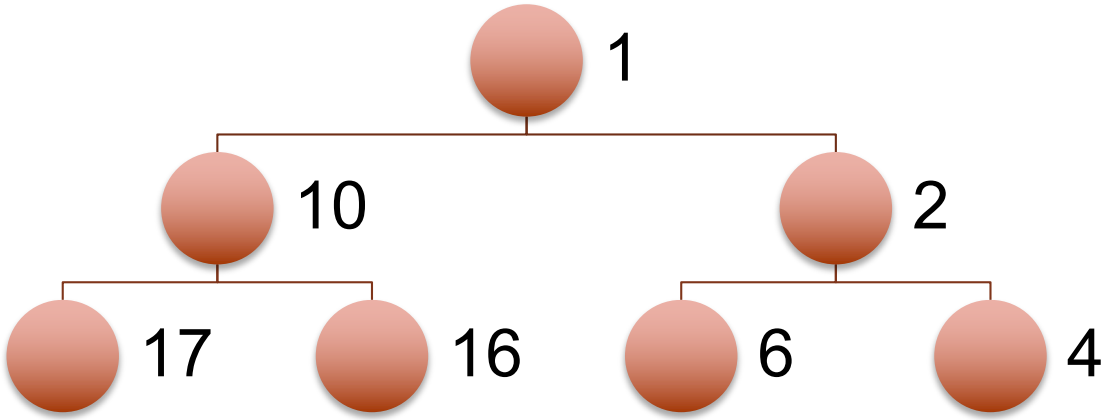
If police officers concentrated on those neighborhoods, they would observe more crimes there.

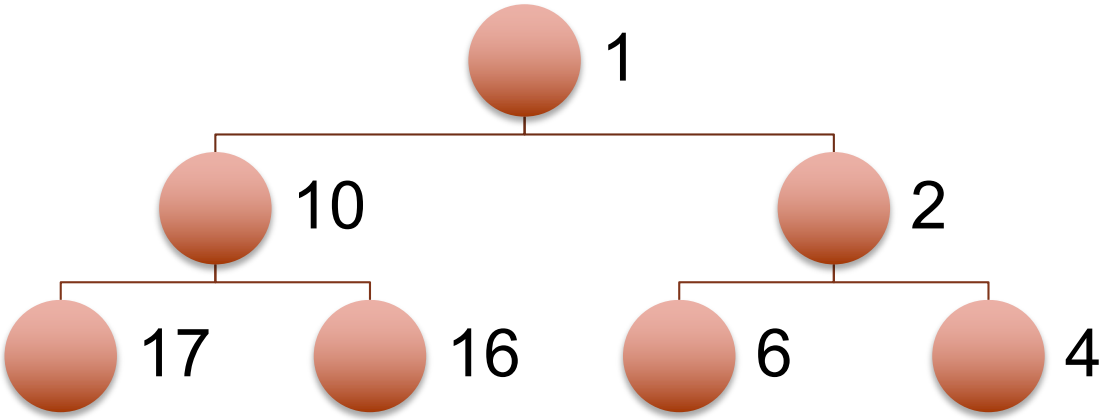
Thus classification as a "high-crime neighborhood" can become a self-fulfilling prophecy or a looping kind.



Predictive policing "hotspot" map

Relevant podcast: https://hiphination.org/season-3-episodes/s3-episode-1-the-precrime-unit/


1
10
2
17
16
6
4



Prioritization and Ranking … A Very, Very, Very Old Tool

Using numbers to categorize people and places and then creating a ranking of those numbers for use in distributing goods and opportunities is as old as the oldest bureaucracies. But computer-based tools make it easier to create, store, share, and act on rankings.

None of this is new … but it is getting faster and easier

The practice of predictive policing based on data about past crime locations goes back to at least the “pin maps” of the 1910s.

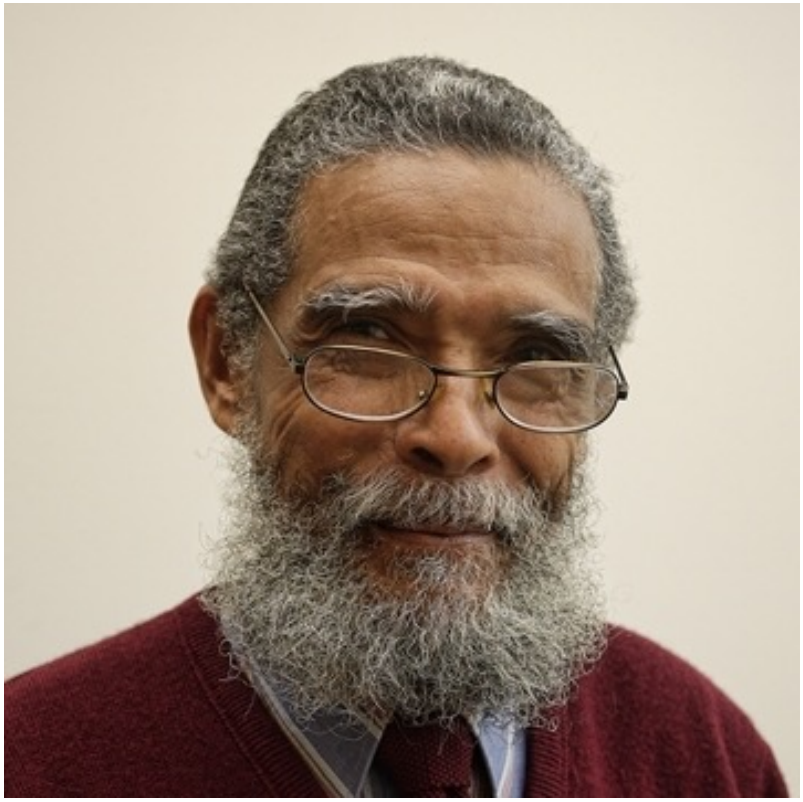
Disproportionate police presence in minority and low-income communities is also not new.

What’s new is the scale of data collection and the ease of algorithmic sorting and prediction thanks to machine learning …

The “new map room” in Scotland Yard, 1947

None of this is new … but it is getting faster and easier

… but even algorithmic decision-making using machine learning is a continuation of the computerized sorting mechanisms based on massive data collection that Oscar Gandy Jr. described in “The Panoptic Sort” (1993)



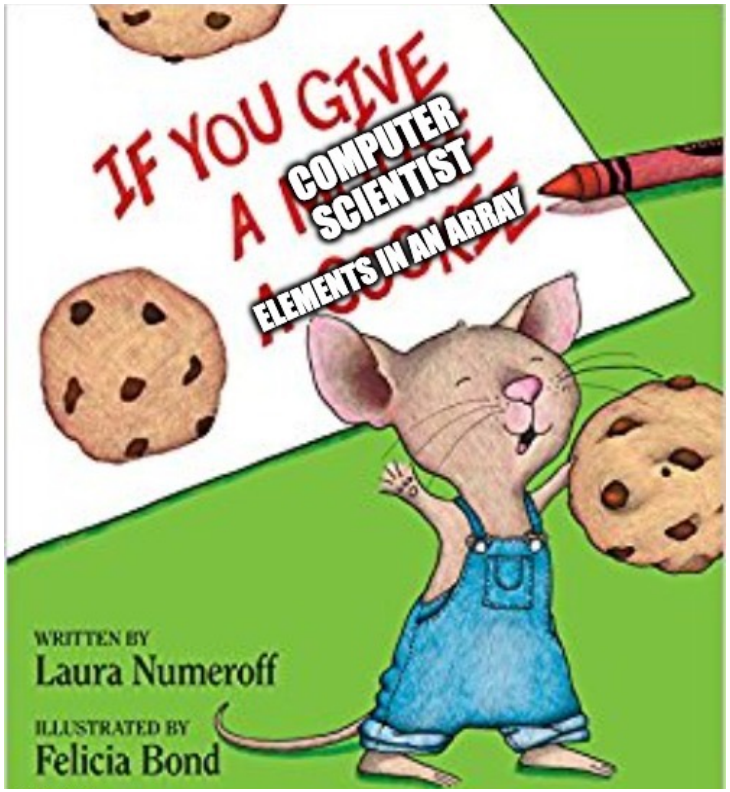
“Read my book! It was re-issued in 2021” – Oscar Gandy Jr. probably

NUMBERS, BINS, AND RANKINGS

Classification & Categories: Where do the Numbers Come From?

In order to create a priority queue, we need to know how to sort the elements … and therefore how to compare them.

```
void sort(Vector<int>& vec) {
  int n = vec.size();
  // already-fully-sorted section grows
  // 1 at a time from left to right
  for (int lh = 0; lh < n; lh++) {
    int rh = lh;
    // find the min element in the
    // entire unsorted section
    for (int i = lh + 1; i < n; i++) {
      // found new min?
      if (vec[i] < vec[rh]) rh = i;
    }
    // swap min into sorted section
    int tmp = vec[lh];
    vec[lh] = vec[rh];
    vec[rh] = tmp;
  }
}
```



… they will probably ask for a way to compare them.

Numbers and letters both have built in orderings. But what about sorting more complex strings?

Names and Alphabetization

Strings might seem easy to alphabetize. But what if the strings store names?

https://tinyurl.com/falsehoodsaboutnames

With the person sitting next to you, look at Falsehoods Programmers Believe about Names, linked above. Which of these would make it difficult to sort names alphabetically for our priority queue?

Hint: check out #14 and #18. Can you find another falsehood about names that would be relevant for ordering them with a priority queue?

See also “Falsehoods programmers believe about …” time, addresses, etc

Two types of harm from classificatory systems

Representational Harm

Harm that arises from a discriminatory or stereotyped representation, or a failure to represent or describe people in the category terms they would use to describe themselves.

Failure to properly represent and alphabetize a name could be a representational harm.

Two types of harm from classificatory systems

Representational Harm

Harm that arises from a discriminatory or stereotyped representation, or a failure to represent or describe people in the category terms they would use to describe themselves.

Failure to properly represent and alphabetize a name could be a representational harm.

Allocative/Distributive Harm

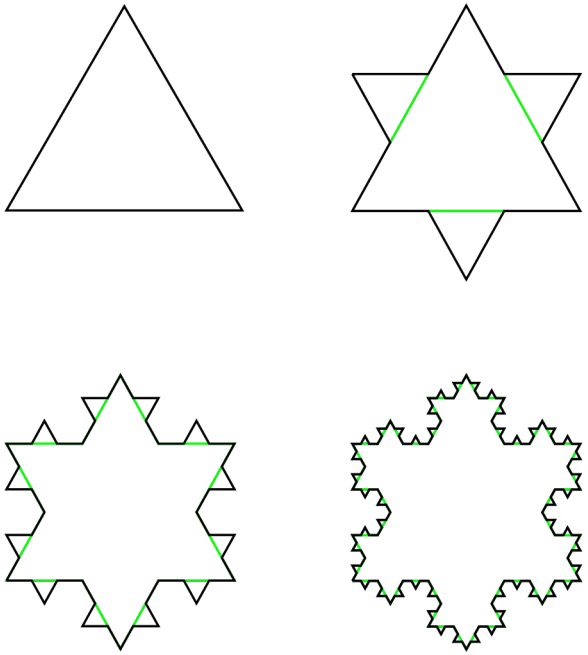
Harm that arises from the unjust distribution of opportunities, resources, goods, or life chances.

Failure to properly represent and alphabetize a name *could* lead to an allocative harm in some circumstances.

Buckets, Bins, and the Granularity of Sorting

Stanford University

Buckets, Bins, and the Granularity of Sorting



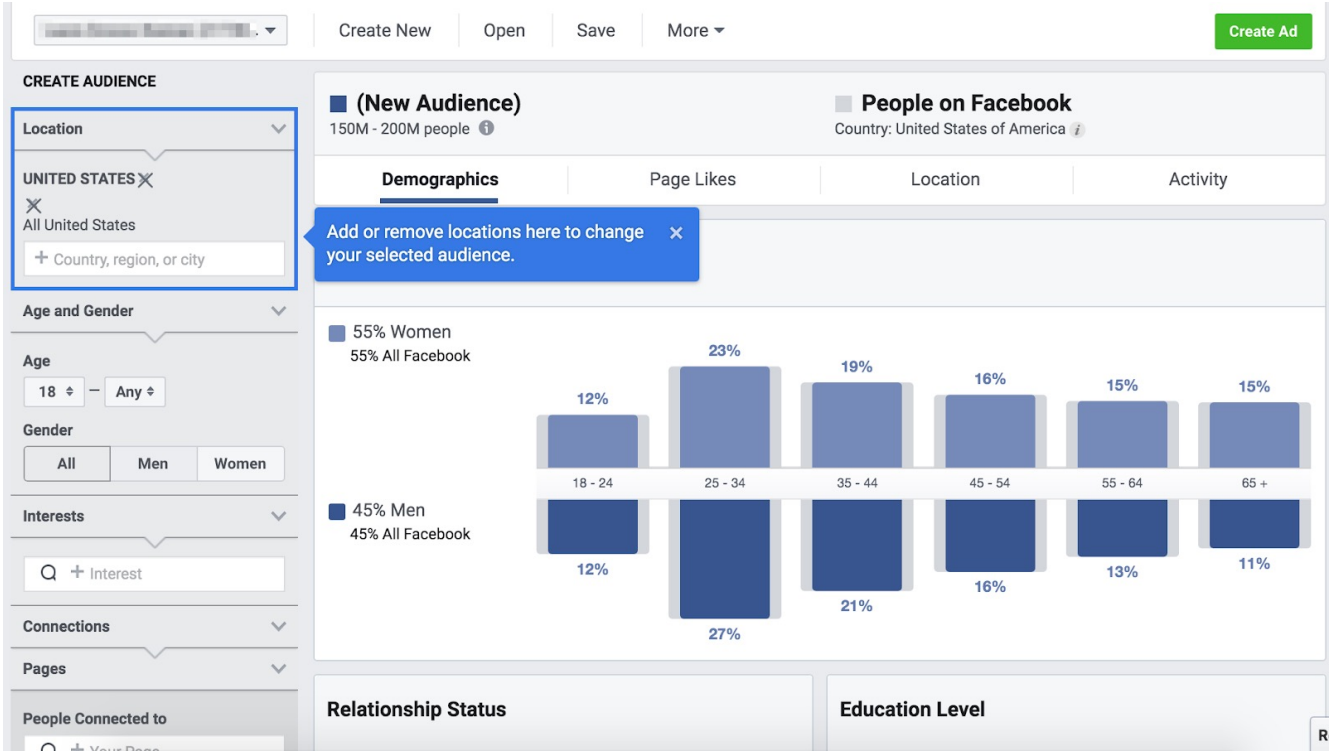
Sorting, Classification, and Gender

Many forms ask for gender, and some only give two options, or two options and “Other.”

Representational Harms in Classificatory Systems

Facebook currently has 14 drop-down gender options. This would be far from covering everyone. Increasing the number of static bins is not enough on its own.

More important is the “Custom” text box. This allows people to display their gender on their profile (although pronoun options are still limited).



However, advertisers see something different

In “audience insights,” the back end shown to advertisers, people are still sorted into women and men only!

On the back end, Facebook (and other companies that rely on advertisements) stores user information & the categories into which they sort users in a data structure (to simplify) and sorts them based on predictions of how likely they are to respond to certain kinds of ads.

Why might this incentivize a company to create fewer rather than more bins? Or to ignore the user's own gender ascription?

What can we do?

Contextual Decision-Making

- Research the complexity of a domain with the help of people who will be affected by the algorithm and other experts.
- Make design choices based on local context
- Expect to revise key design elements (rankings, binnings) over time as the context shifts

What can we do?

Contextual Decision-Making

- Research the complexity of a domain with the help of people who will be affected by the algorithm and other experts.
- Make design choices based on local context
- Expect to revise key design elements (rankings, binnings) over time as the context shifts

Participatory Design (or co-design)

- Work with people classified, ranked, and prioritized by your system to negotiate categories that are meaningful to them
- Participants help define the problem to be solved and to evaluate proposed solutions
- Participants embody the fractal diversity imperfectly captured by your classification scheme – determine with them which groupings are acceptable.

Thank you!

KATIE CREEL – KCREEL@STANFORD.EDU

SET UP A MEETING ANY TIME AT

CALENDLY.COM/KATHLEENCREEL