

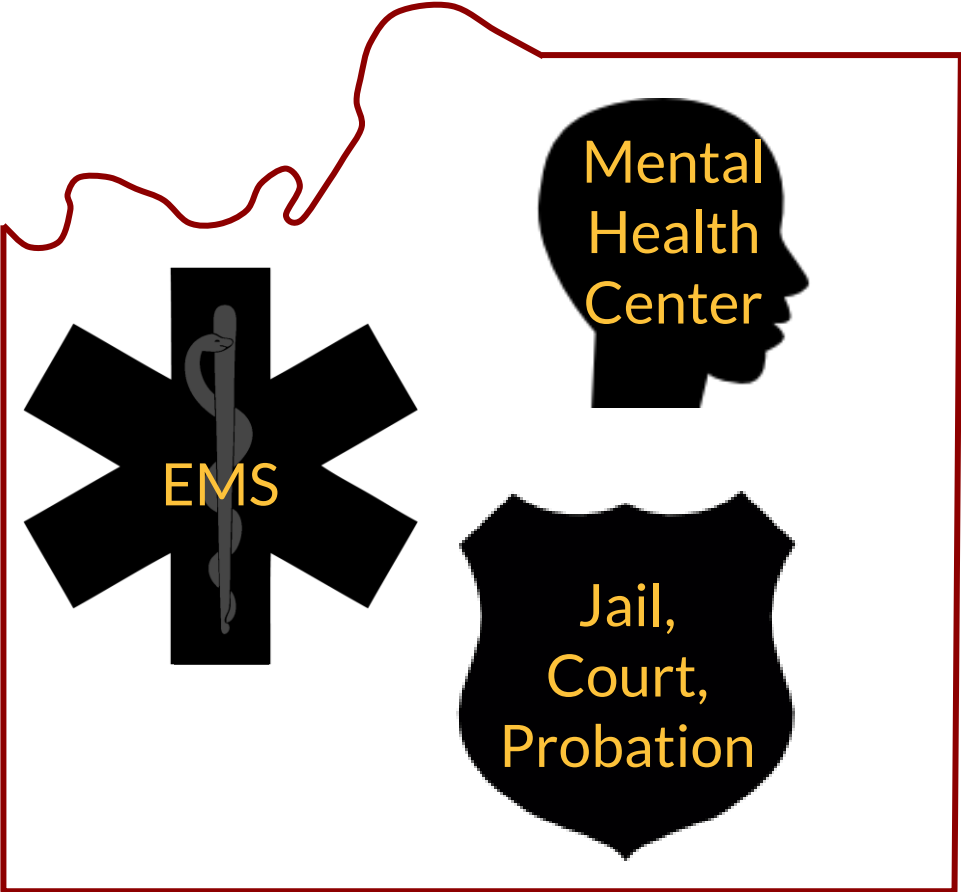
Using Data for Cross-Systems Resource Targeting

June 5, 2017

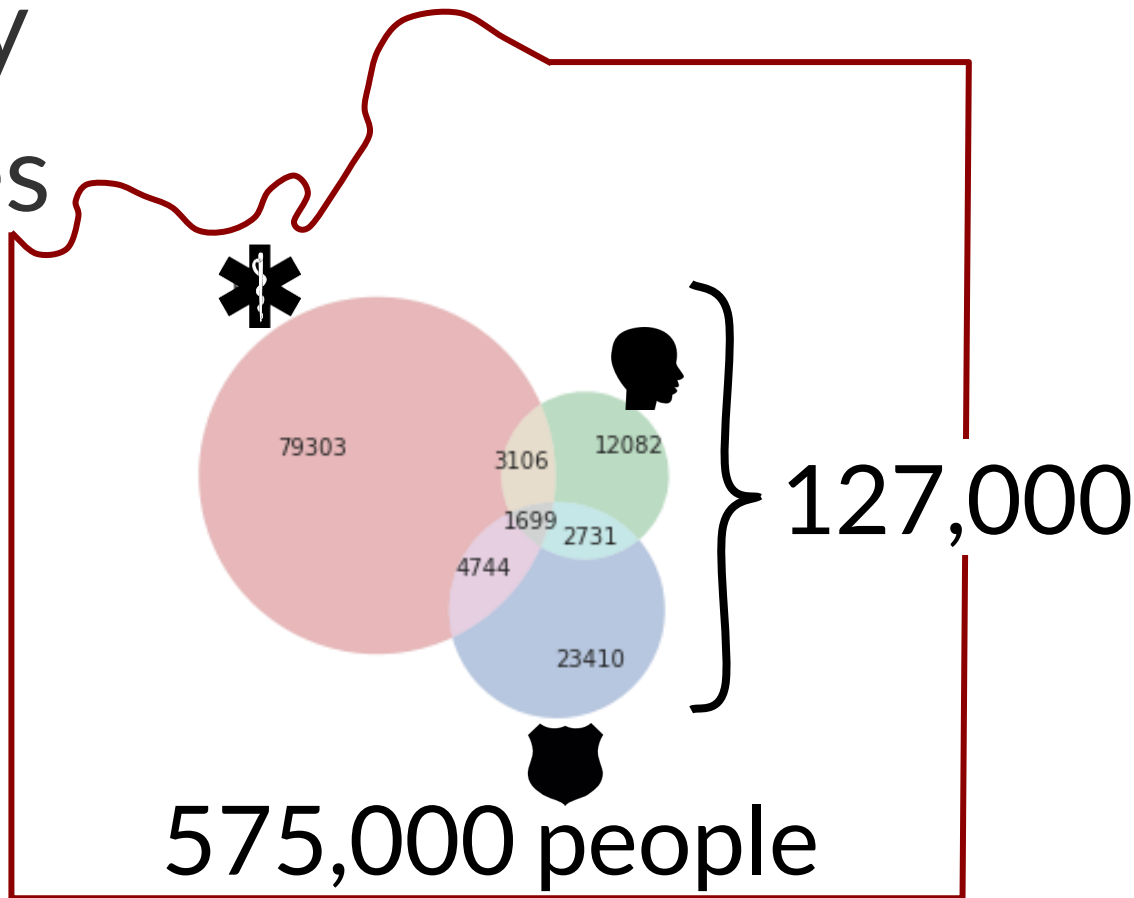
Center for Data Science & Public Policy



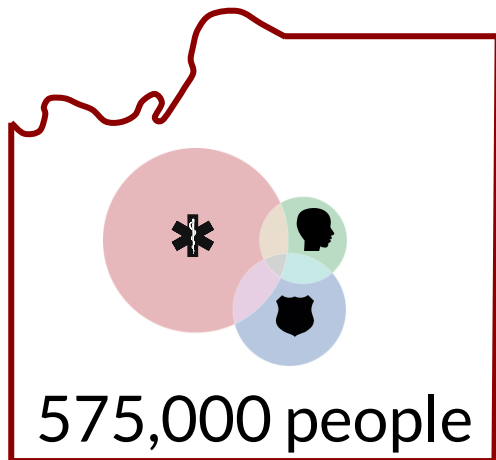
Johnson County Services



County Services



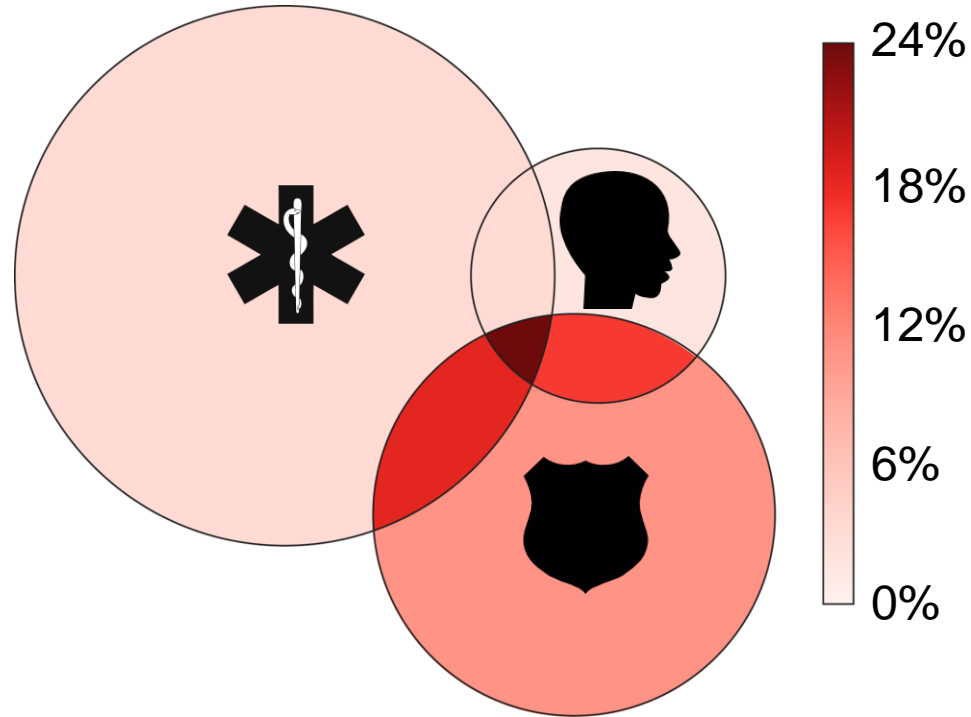
Key data requirements



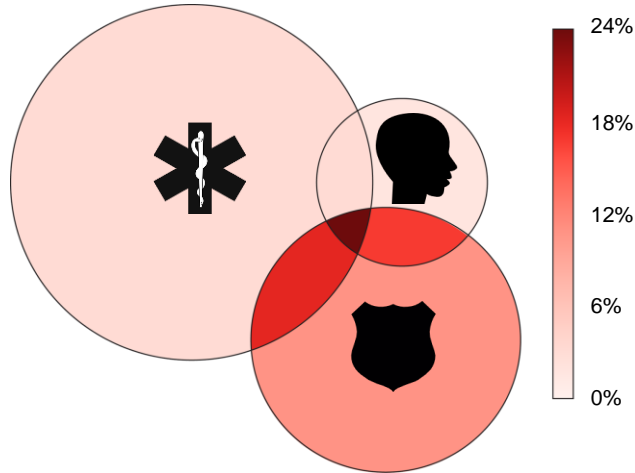
Venn diagram

- Consistent combinations of identifying information:
 - Name fields
 - Date of birth
 - SSN
 - Gender
 - Race
 - Address
- Data cover similar time periods
- Additionally requires consistent collection and granularity of residence address to normalize by a total population

Percent of population that entered jail in 2015



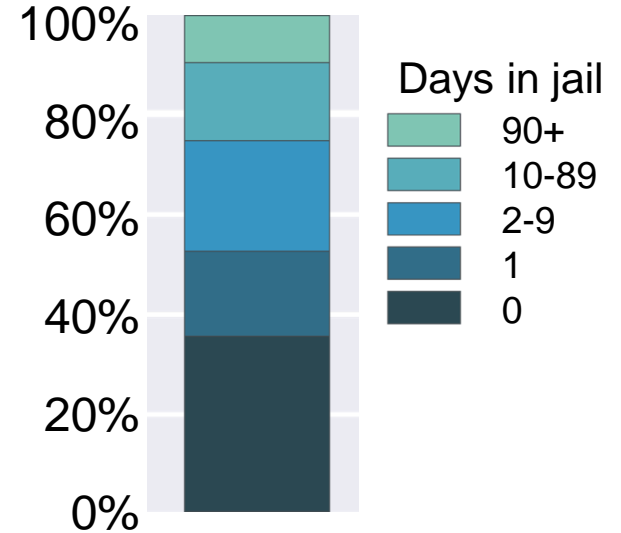
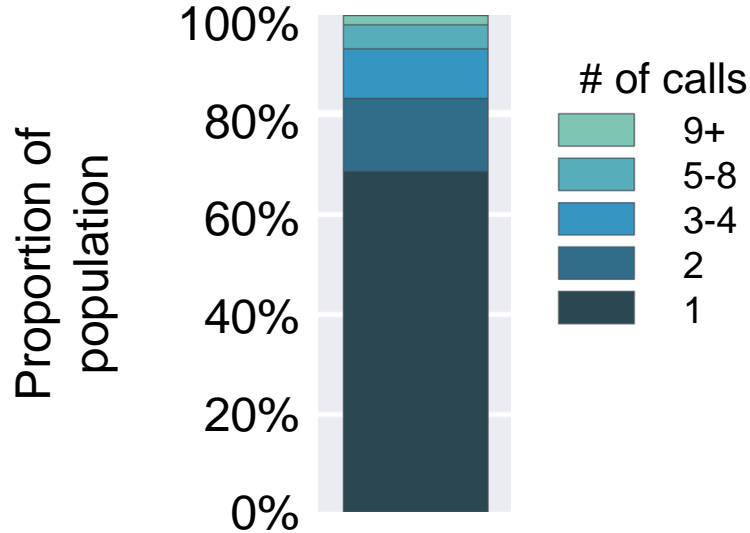
Key data requirements



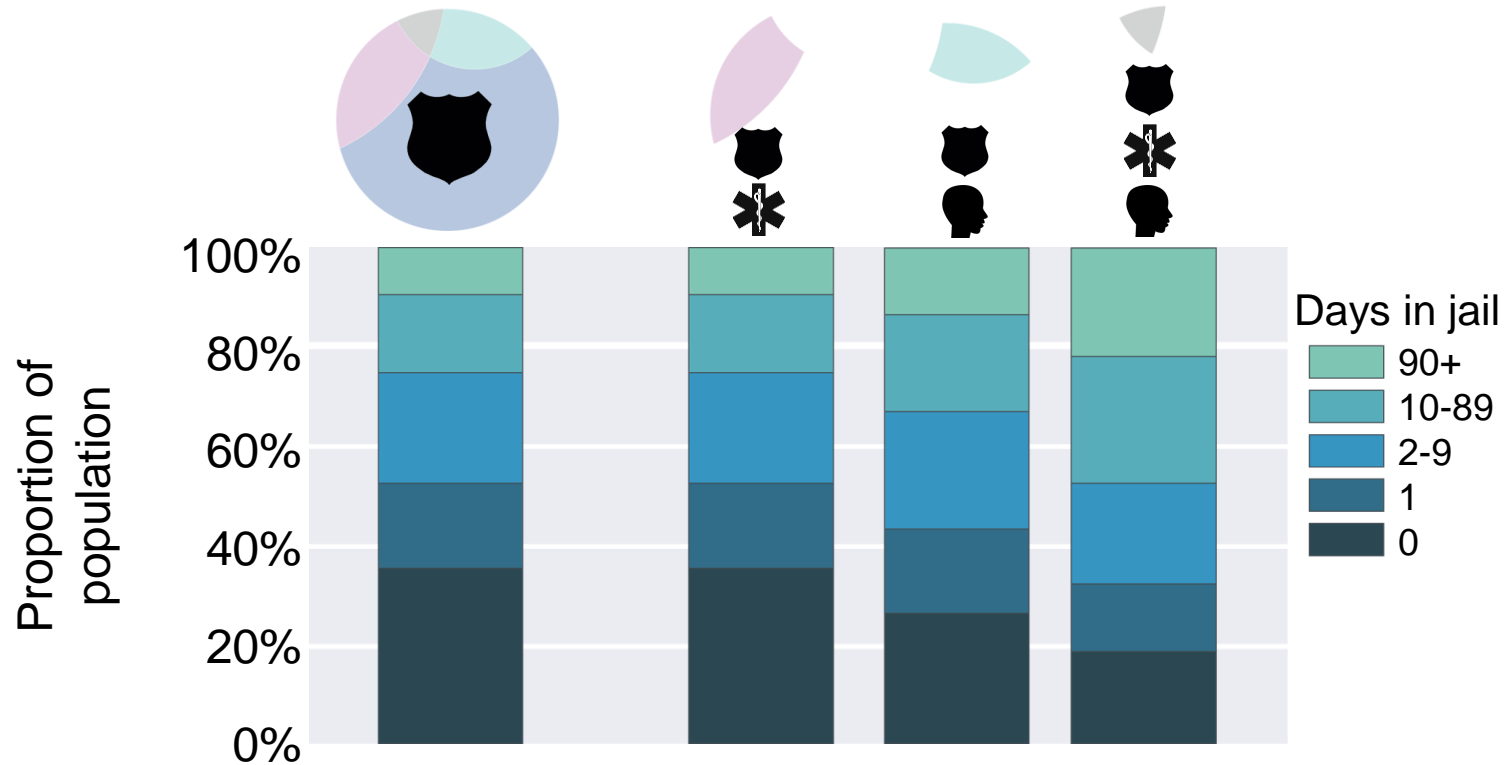
Risk comparisons

- Everything required for Venn diagram, plus...
- One record per interaction
- Date of interaction

“Frequent-flyer” explorations



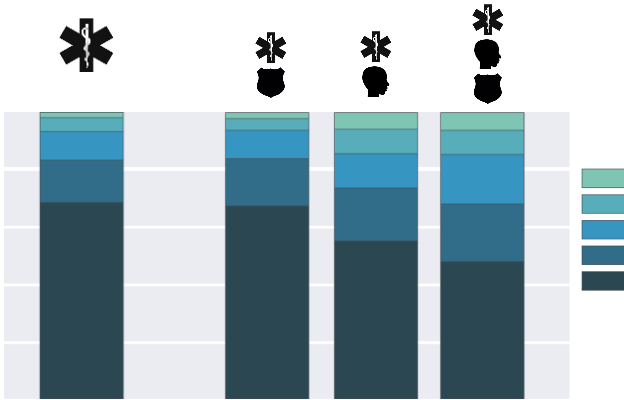
Prolonged interactions: Jail



Key data requirements

Cross-silo frequencies

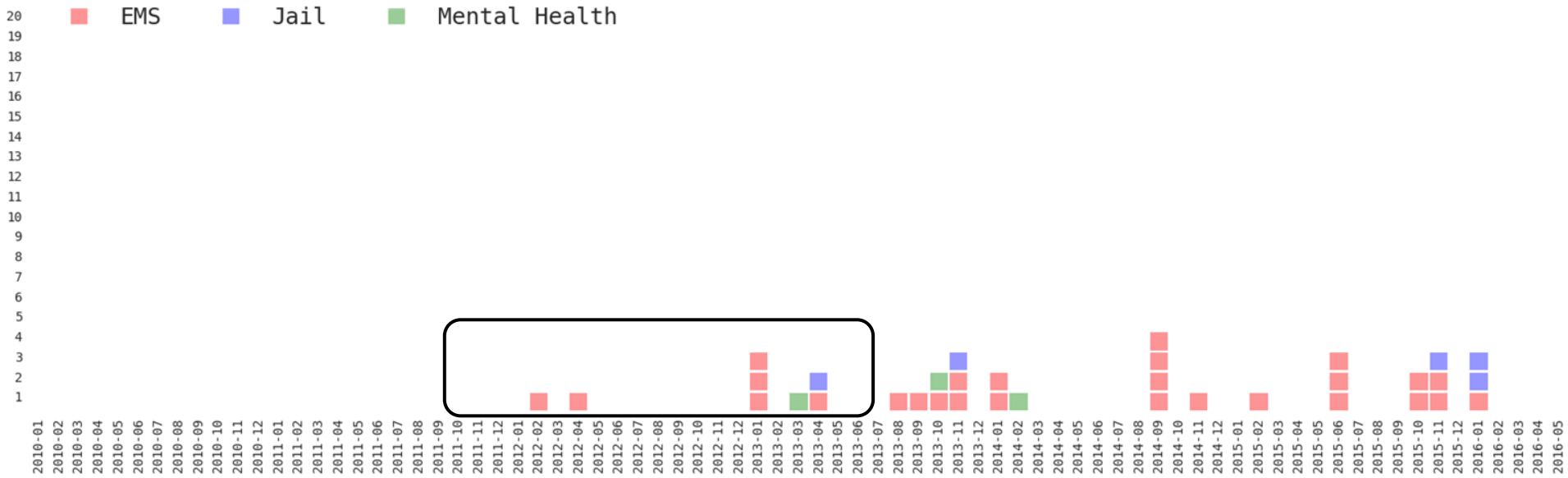
- Everything required for Venn diagram and risks, plus...
- Interaction type
- Duration of interaction



Sequence Analysis

Service Path Order:

E → E → E → E → E → M → E → J



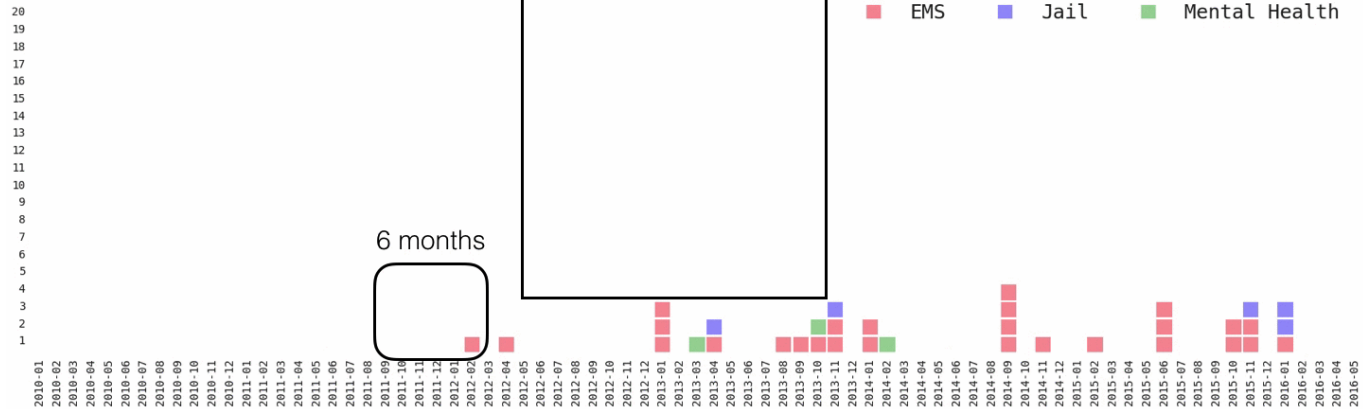
Generalized Sequential Pattern

Specify time window size

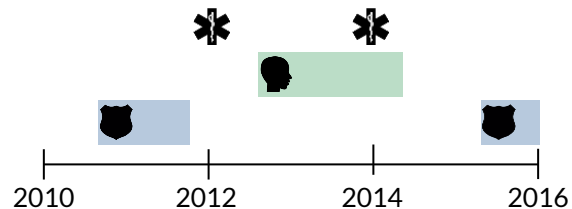
2 years, 1 years, 6 months, 3 months, 1 weeks, 3 days, etc.

Generate subsequence dataset

Subsequence dataset



Key data requirements



Sequence features

- Crucial to incorporate both start and stop dates of service/booking
- Note distinction between case management start/stop and interaction start/stop
- Documented reason of interaction end is valuable (good behavior, program drop out, administrative reasons, death)



Demographics

Age at earliest interaction with a public system

Age group at last interaction with a public service



Counts of Interactions

Number of bookings in last year

Number of mental health entries in the last year

Total number of bookings



Interaction Context

Number of therapists seen

Number of mental health services used

Type of therapy

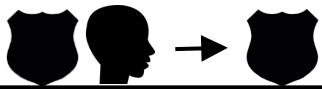
Average bail amount



Timeline

Standard deviation of time between public system interaction

Had two bookings within a year



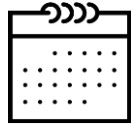
Matt Bauman	6.94
Eddie Lin	6.17
Kate Boxer	5.79
Erika Salomon	5.75
Lauren Haynes	5.02
Joe Walsh	4.72
Jen Helsby	4.49
Rayid Ghani	4.28
Steve Yoder	3.85
Chris Schnewies	3.64
Robert Sullivan	3.51

Prioritized List: top 200 people

Matt Bauman	6.94
Eddie Lin	6.17
Kate Boxer	5.79
Erika Salomon	5.75
Lauren Haynes	5.02
Joe Walsh	4.72
Jen Helsby	4.49
Rayid Ghani	4.28
Steve Yoder	3.85
Chris Schnewies	3.64
Robert Sullivan	3.51



102 individuals



19 years total jail time



\$250,000 absolute minimum cost



2 years since last mental health contact

Key data requirements



Matt Bauman	6.94
Eddie Lin	6.17
Kate Boxer	5.79
Erika Salomon	5.75
Lauren Haynes	5.02
Joe Walsh	4.72
Jen Helsby	4.49
Rayid Ghani	4.28
Steve Yoder	3.85
Chris Schnewies	3.64
Robert Sullivan	3.51

Predictive analytics

- **ALL** available data from each silo
- We used over 300 indicators for each individual that were generated from the richness of the datasets of each silo:
 - Bail amounts
 - Type of therapy
 - Number of therapists
 - EMS primary impression
- Several (>3) years of history for all silos
- Existing baselines (thresholds) used to identify super-utilizers

Note:

- You can use the more robust “predictive” analysis to set more naive thresholds