Redirecting People with Complex Conditions to Effective Care

October 6, 2016

Center for Data Science & Public Policy



dsapp.uchicago.edu

Center for Data Science and Public Policy University of Chicago



Center for Data Science and Public Policy University of Chicago



Center for Data Science and Public Policy University of Chicago



Center for Data Science and Public Policy University of Chicago



Center for Data Science and Public Policy University of Chicago

Percent of population that entered jail in 2015



dsapp.uchicago.edu

Center for Data Science and Public Policy University of Chicago



Center for Data Science and Public Policy University of Chicago

Sequence Analysis

Service Path Order: $E \rightarrow E \rightarrow E \rightarrow E \rightarrow E \rightarrow M \rightarrow E \rightarrow J$



dsapp.uchicago.edu

Center for Data Science and Public Policy University of Chicago

Generalized Sequential Pattern



Center for Data Science and Public Policy University of Chicago

dsapp.uchicago.edu



Center for Data Science and Public Policy University of Chicago



Age at earliest interaction with a public system Age group at last interaction with a public service



Number of bookings in last year Number of mental health entries in the last year Total number of bookings



Number of therapists seen Number of mental health services used Type of therapy Average bail amount



Standard deviation of time between public system interaction Had two bookings within a year

Center for Data Science and Public Policy University of Chicago

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

Center for Data Science and Public Policy University of Chicago

dsapp.uchicago.edu

Johnson County : Outcomes

dsapp.uchicago.edu

Center for Data Science and Public Policy University of Chicago



Center for Data Science and Public Policy University of Chicago

Prioritized List: top 200 people

52% (104) of the (top) 200 predicted individuals end up going to Jail in the next year

dsapp.uchicago.edu

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









2 years since last mental health contact

Center for Data Science and Public Policy University of Chicago

Joco: How we got there

- Data Use agreements internally & Consolidation of data
- Co-responders
- Matching
- Anonymization

SLC: How we got there

- Data Use agreements internally & Consolidation of data
- Sharing data with UChicago



	Health Systems	Criminal Justice	Behav Hea	rioral Ith
OHNSON COUNTY KANSAS	Paramedic transport logs NOT: EMTs, ERs	Jail bookings Court records Probation records Not: arrests, dispatches	County mental health case management - Diagnoses - Services - Discharges	
SALT LAKE COUNTY		Jail bookings Court records Probation records Not: arrests, dispatches	Mental health billings - Services - Dates	Substance abuse treatment (TEDS)

Center for Data Science and Public Policy University of Chicago



Venn diagram

- Requires linking individuals across multiple systems
 - Consistent combinations of identifying information can be used to match afterwards (more is better, must be normalized across silos):
 - Name fields
 - Date of birth
 - SSN
 - Gender
 - Race
 - Address
- Important to ensure data covers similar time periods



Venn diagram

- Requires linking individuals across multiple systems
 - Consistent combinations of identifying information can be used to match afterwards (more is better, must be normalized across silos):
 - Name fields
 - Date of birth
 - SSN
 - Gender
 - Race
 - Address
- Important to ensure data covers similar time periods
- Additionally requires consistent collection and granularity of residence address to normalize by a total population



^oroportion of

Within-silo frequency

- One record per interaction
- Interaction type
- Consistent patient identifiers across interactions
 - Bonus points for linking repeated visits to the same ID at time of interaction
 - Otherwise, consistent combinations of identifying information can be used to match afterwards (more is better):
 - Name fields
 - Date of birth
 - SSN
 - Gender
 - Race
 - Address

Center for Data Science and Public Policy University of Chicago



Across-silo frequencies

- One record per interaction
- Interaction type
- Date of interaction
- Duration of interaction
- Requires linking/matching from Venn diagram

dsapp.uchicago.edu

Center for Data Science and Public Policy University of Chicago



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)

Center for Data Science and Public Policy University of Chicago

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 superutilizers

Note:

• You can use the more robust "predictive" analysis to set more naive thresholds

Center for Data Science and Public Policy University of Chicago

dsapp.uchicago.edu

Next Steps

- Asking for new jurisdictions
 - Data Use Agreements, Data & Data Transfer
 - Funding
- Tech Consortium Work
 - Open platform with the de-identified data to expand access to other researchers & analysts

Questions?

Contact :

- UChicago : Lauren Haynes (Lnhaynes@uchicago.edu)
- Johnson County :
 - Robert Sullivan (Robert.Sullivan@jocogov.org)
 - Chris Schneweis (Chris.Schneweis@jocogov.org)
 - Steve Yoder (Steve.Yoder@jocogov.org)
- Salt Lake County :
 - Fraser Nelson (Fnelson@slco.org)

Appendix



Center for Data Science and Public Policy University of Chicago

Removing Identifying Information: hashing

One-way only

Lauren Haynes	e309c1ba03b22b72bc46cdf4200e0d19
Jane Doe	73c01bf88feb18695bd65e611ef1cf26
Matthew Bauman	18750ae79e3e94df96fdd4a354dbb2b0

655-82-8799	dfe1f020f8b457792a628692a607f94e
999-99-9999	007e3c1b311ec3848c878a420736c09f
000-00-0000	072f1bbdf1984fc0988be2d4b0c91803

d	sa	pr).u	ch	icag	o.ec	du

Center for Data Science and Public Policy University of Chicago



Center for Data Science and Public Policy University of Chicago

Sequence Analysis

Service Path Order: $E \rightarrow E \rightarrow E \rightarrow E \rightarrow E \rightarrow M \rightarrow E \rightarrow J$



dsapp.uchicago.edu

Center for Data Science and Public Policy University of Chicago

Generalized Sequential Pattern



Center for Data Science and Public Policy University of Chicago

dsapp.uchicago.edu



Center for Data Science and Public Policy University of Chicago

Generalized Sequential Pattern

$\textcircled{E} \Rightarrow E \Rightarrow E \Rightarrow E \Rightarrow E \Rightarrow M \Rightarrow E \Rightarrow J$

dsapp.uchicago.edu

Center for Data Science and Public Policy University of Chicago

Generalized Sequential Pattern

in 6 months Specify time window size E→E E→E F→J J→J Generate subsequence dataset E→M→J E→E→M $E \rightarrow E \rightarrow E \rightarrow E$ J→E $E \rightarrow E \rightarrow E \rightarrow E \rightarrow E$ $E \rightarrow E \rightarrow E$ **Break into subsets** 📊 E->E->E->E 🏡 📅 E->E->M 🍫 📅 E->M->J 🍇 1 0 9 0 Sequential pattern mining 10 0 0 11 0 12 0 1 0 13 0

Frequent Closed Sequence Mining without Candidate Maintenance, J. Wang, J. Han, and C. Li, IEEE Trans. on Knowledge and Data Engineering 19(8):1042-1056, IEEE Press, Piscataway, NJ, USA 2007

Top-K Frequent Patterns

Center for Data Science and Public Policy University of Chicago

dsapp.uchicago.edu

Choose patterns to be features

Figure 2. The four most important predictors for officers flagged by DSaPP's system.

OFFICER 1	High number of rule of conduct violations in last 15 years	Officer was suspended in last 15 years	High number of counseling interventions after special investigations	High number of sustained complaints in last 15 years
OFFICER 2	High number of counseling interventions after special investigations	High number of rule of conduct violations in last 15 years	High number of rule of prior adverse incidents in last 15 years	High number of special investigations correctives written in last 15 years
OFFICER 3	High number of complaints against officer in last 15 years	High number of rule of conduct violations in last 15 years	Officer was suspended in last 15 years	High number of counseling interventions after special investigations
OFFICER 4	Officer has dealt with high number of domestic violence incidents	High number of special investigations correctives written in last 15 years	Officer was suspended in last year	High number of accidents in last year
35	Officer has dealt	High number		Officer was

dsap

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

Anonymization, Hashing, Matching and Privacy

- The need for individual-level, linkable data
 - Anonymization Before Matching vs Anonymization after Matching
- Flow less sensitive information into more sensitive areas (Jail -> Mental Health)
- Hash anything that is otherwise public record (KDOC numbers) that would de-anonymize your data set
- Need a process owner for release of data (specific events that make the news can be de-anonymized)

Removing Identifying Information: Hashing

