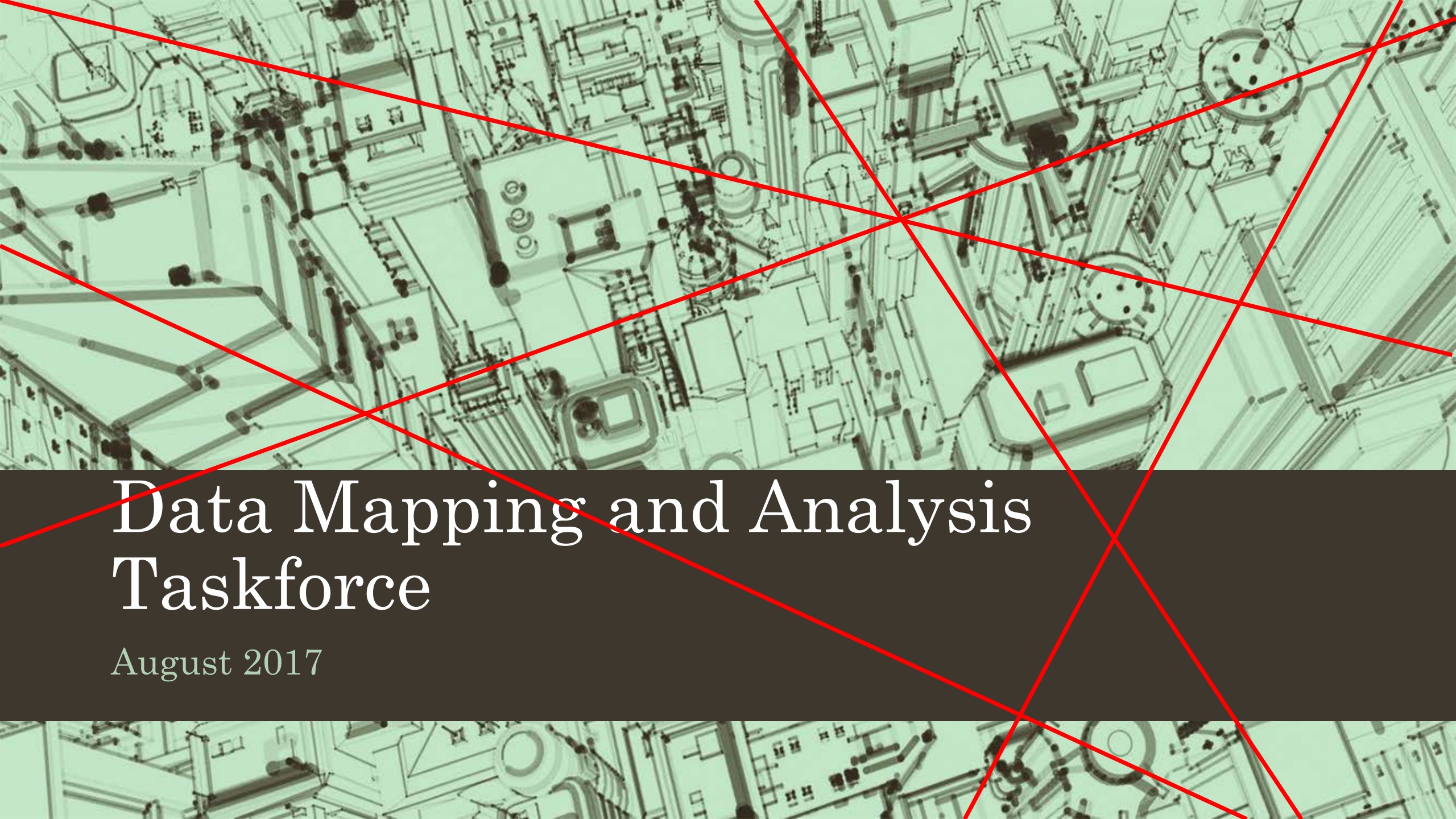


Data Mapping and Analysis Taskforce

August 2017

Importance of Data:

- Where we live, Where we go, what we buy, what we say.
- It is being compiled, but there is a trace in several different sources
- Active Measurement produced data
- We measure to improve
- More Data we get the bigger problems we can solve
- Visualizing data allows us to see how complex systems function.

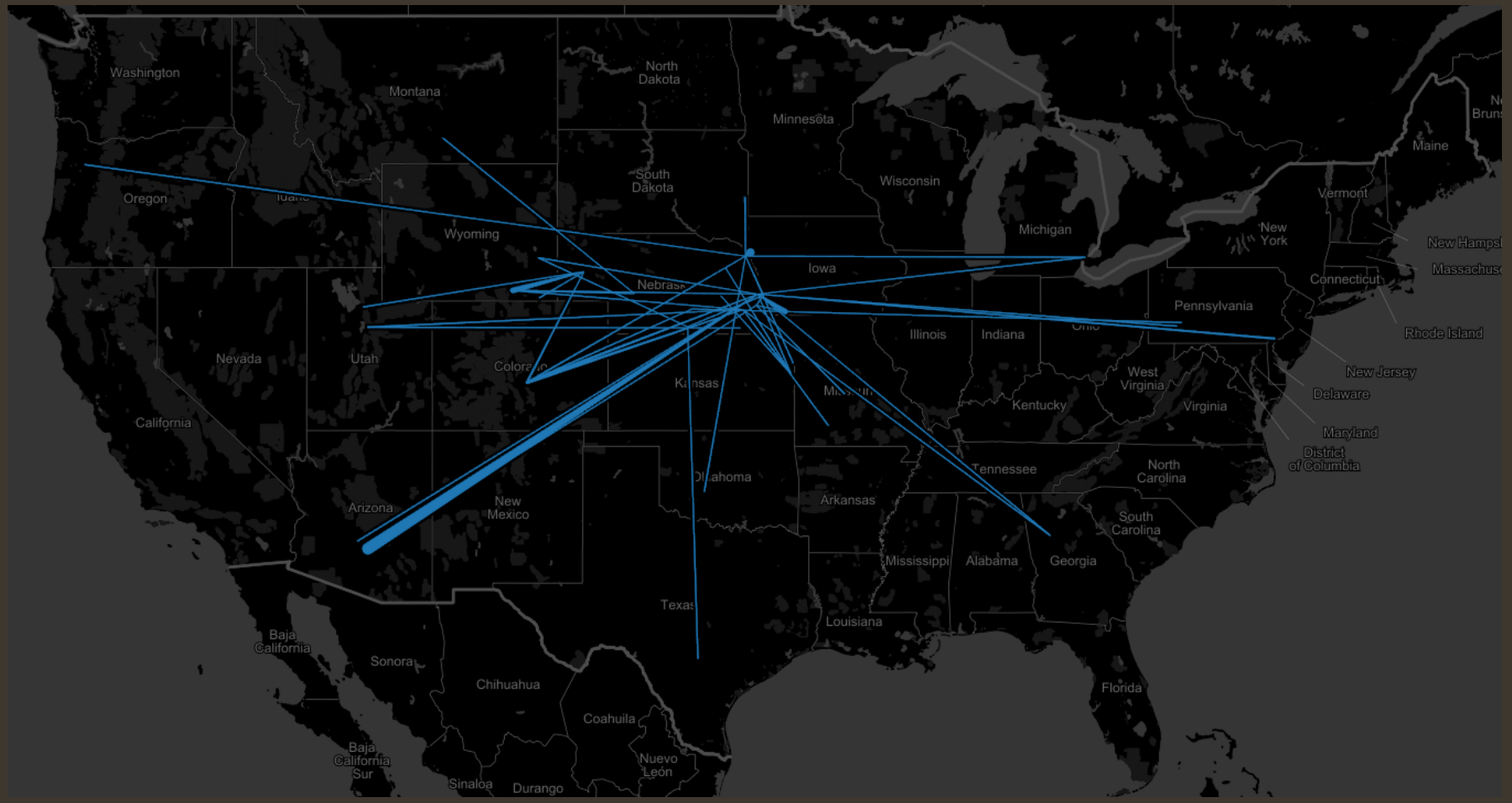
Taskforce Goal

- The current scope of the taskforce is to better understand the proximity between a youth's placement and their residence and if there is a way to use existing facilities in order to pilot a multi-level of care system.
- To answer these questions, the DMA Taskforce first investigated the proximity of out-of-state probation placements and placements to the YRTCs.
- The goal of the analysis is to inform stakeholders of the distance between a youth's placement and their residence

- Amanda Felton (Resource)
- Anne Hobbs
- Bethany Allen (Resource)
- Jana Peterson
- Juliet Summers
- Katherine Bass
- Mike Fargen (Chair)
- Monica Miles-Steffens

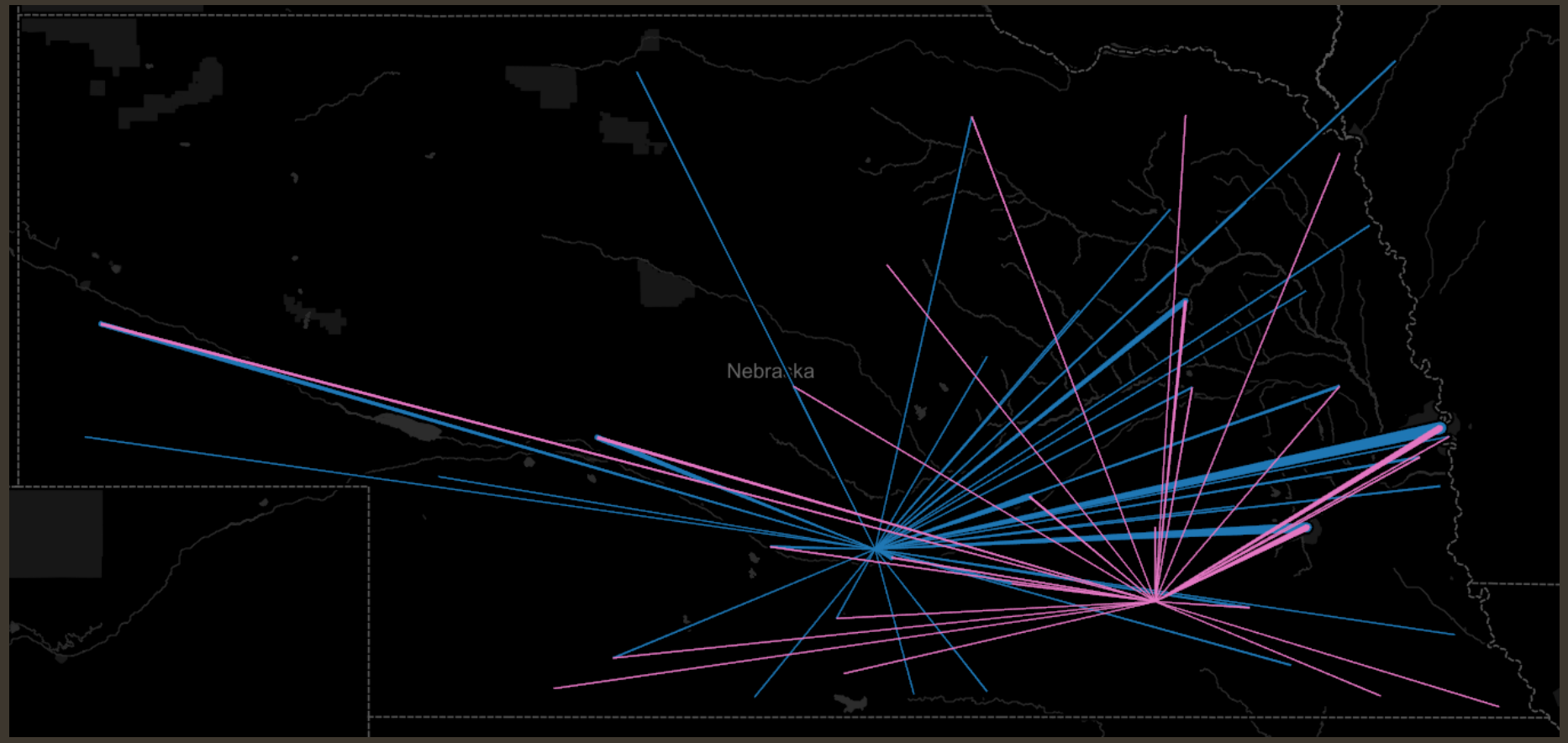
Preliminary Results (Out-of-State Probation Population)

- 11 Months of Data
- 144 Records
- 469.7 Average Estimated Distance
- 30.6% of Population within 120 miles



Preliminary Results (YRTC Population)

- 23 Months of Data
- 315 Records
  - 220 Male
  - 95 Female
- Avg. Est. Distance:
  - Male = 121.1 m
  - Female = 108.3 m
- % Within 120 miles
  - Male = 45.0%
  - Female = 77.9%

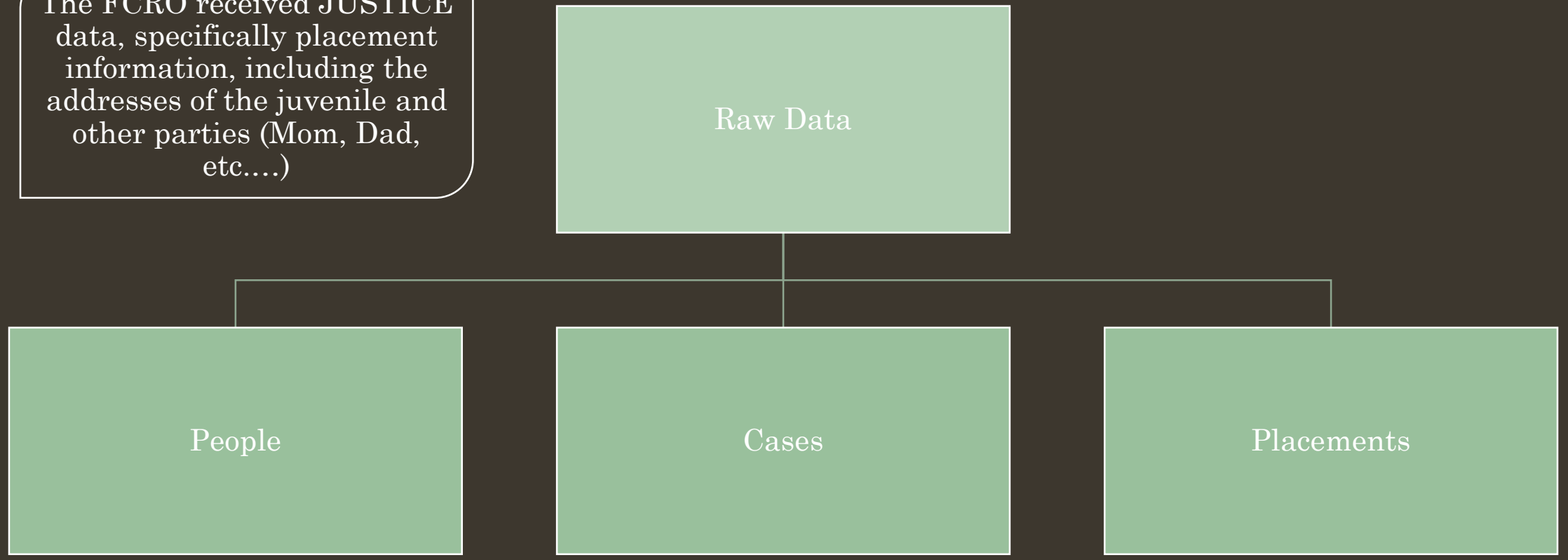


Mapping the Cost of Justice | The Human Face of Big Data

http://www.pbs.org/show/human-face-big-data/

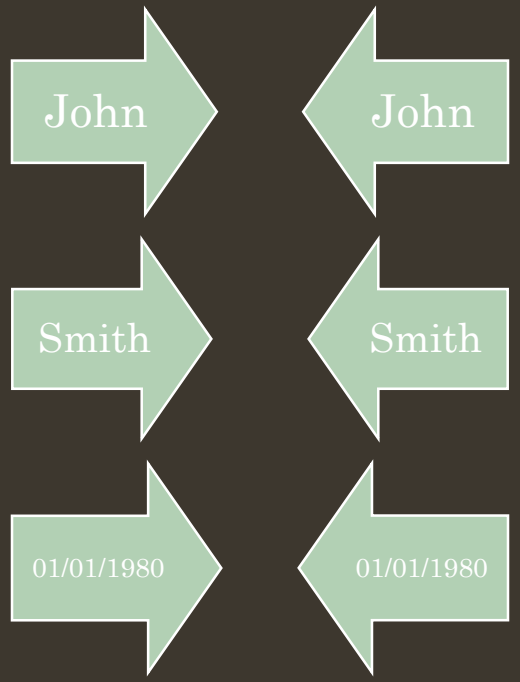
JUSTICE DATA RESHAPING

The FCRO received JUSTICE data, specifically placement information, including the addresses of the juvenile and other parties (Mom, Dad, etc.…)

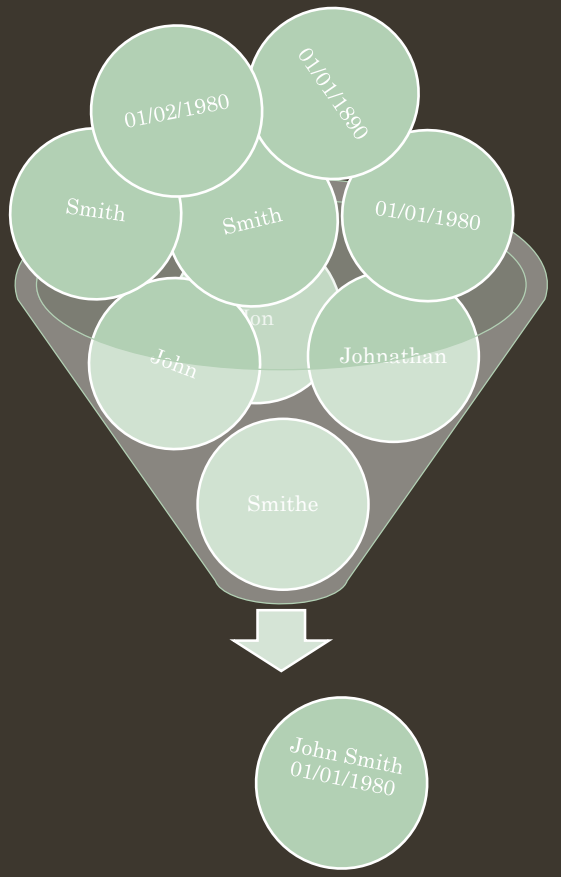


JUSTICE Juvenile Record Linkage

Deterministic



Probabilistic



Probabilistic Record Linkage Software: Link Plus

- Link Plus is a probabilistic record linkage program developed at the U.S. Centers for Disease Control and Prevention (CDC), Cancer Division.
- Link Plus was written as a linkage tool for cancer registries, in support of CDC's National Program of Cancer Registries.
- It is an easy-to-use, stand-alone, Windows application that can be run in two modes:
  - Detect Duplicates
  - Link to Other
- Link Plus provides an option that allows you to use the name frequencies of 1990 Census data or National Death Index data when the current data file specified as File 1 does not provide reliable estimates of the distributions of last name and first name, which is often the case when you are working with small datasets.
- To compute the default M-probabilities, Link Plus uses the data to generate the frequencies of last names and first names and then computes the weights for last name and first name based on the frequencies of their values.

Field	m-prob	u-prob	agree	disagree
First Name	0.96	0.00191	5.66119	-2.92821
Last Name	0.97	0.00102	6.24490	-3.19088
Date of Birth	0.96	0.00069	6.58766	-2.92932

m-prob: The probability that a matching variable agrees given that the comparison pair being examined is a match. The M-probability measures the reliability of each data item. A Value of 0 means the data item is totally unreliable (0%) and a value of 1 means that the data item is completely reliable (100%). Reasonable values range from 0.9 (90% reliable) to 0.9999 (99.99% reliable).

u-prob: The probability that a matching variable agrees given that comparison pair being examined as a non-match

agree: The agreement weight assigned for an agreement on a given matching variable

disagree: The disagreement weight assigned for a disagreement on a given matching variable

JUSTICE Matching Algorithm

- **Jaro-Winkler Metric**
- The Jaro-Winkler Metric is a string comparator which measures the partial agreement between two strings. In many matching situations, it is not possible to compare two strings exactly (character-by-character) because of typographical errors. Dealing with typographical errors via approximate strings comparison has been a major research effort in computer science. Jaro introduced a string comparator that accounts for random insertion, deletions, and transpositions. In a small study, Winkler showed that the Jaro comparator worked better than some other available comparators. In a large study, Budzinsky concluded that the comparators due to Jaro and Winkler were the best among twenty comparators available in computer science literature.
- The basic Jaro algorithm consists of three procedural components: (1) compute the string length, (2) find the number of common characters in the two strings, and (3) find the number of transpositions between the two strings. The definition of common characters used is that any agreeing characters must be within half the length of the shorter string. The definition of transposition is that the character from one string is out of order with the corresponding common character from the other string. Winkler enhanced the Jaro string comparator by assigning increased value to agreement on beginning characters of a string. This enhancement was based on ideas from a very large empirical study by Pollock and Zamora for the Chemical Abstract Service. The study showed that the fewest errors typically occur at the beginning of a string and that error rates by character position increase monotonically as the position moves to the right.
- The formula for the basic Jaro string comparator is as follows:
- The number of transpositions is calculated as follows: The first common character on one string is compared to the first common character on the other string. If the characters are not the same, half of a transposition has occurred. Then the second common character on one string is compared to the second common character on the other string, etc. The number of mismatched characters is divided by two to yield the number of transpositions.

JUSTICE Matching System

- The Soundex system is over 120 years old, and was first applied to 1880 census data. The Soundex code for a name consists of a letter followed by three numbers: the letter is the first letter of the name, and the numbers encode the remaining consonants. Zeroes are added at the end if necessary to produce a four-character code. Additional letters are disregarded.
- Example: Washington is coded W-252 (W, 2 for the S, 5 for the N, 2 for the G (remaining letters disregarded)
- Using the Soundex code phonetic system reduces matching problems due to different spellings, and is simple and fast.

JUSTICE Scored Matching

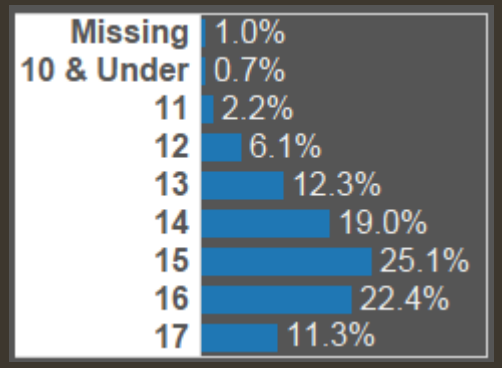
- Cutoff Value < 5.0
- The Cut Off Value is the linkage score For a comparison pair, the overall weight over all matching variables; a higher score means a higher likelihood of being a match. value above which comparison pairs are accepted as potential links. Enter a value in the box provided. The value should always be positive.
- Work Down
- Work Up
- Manual Review

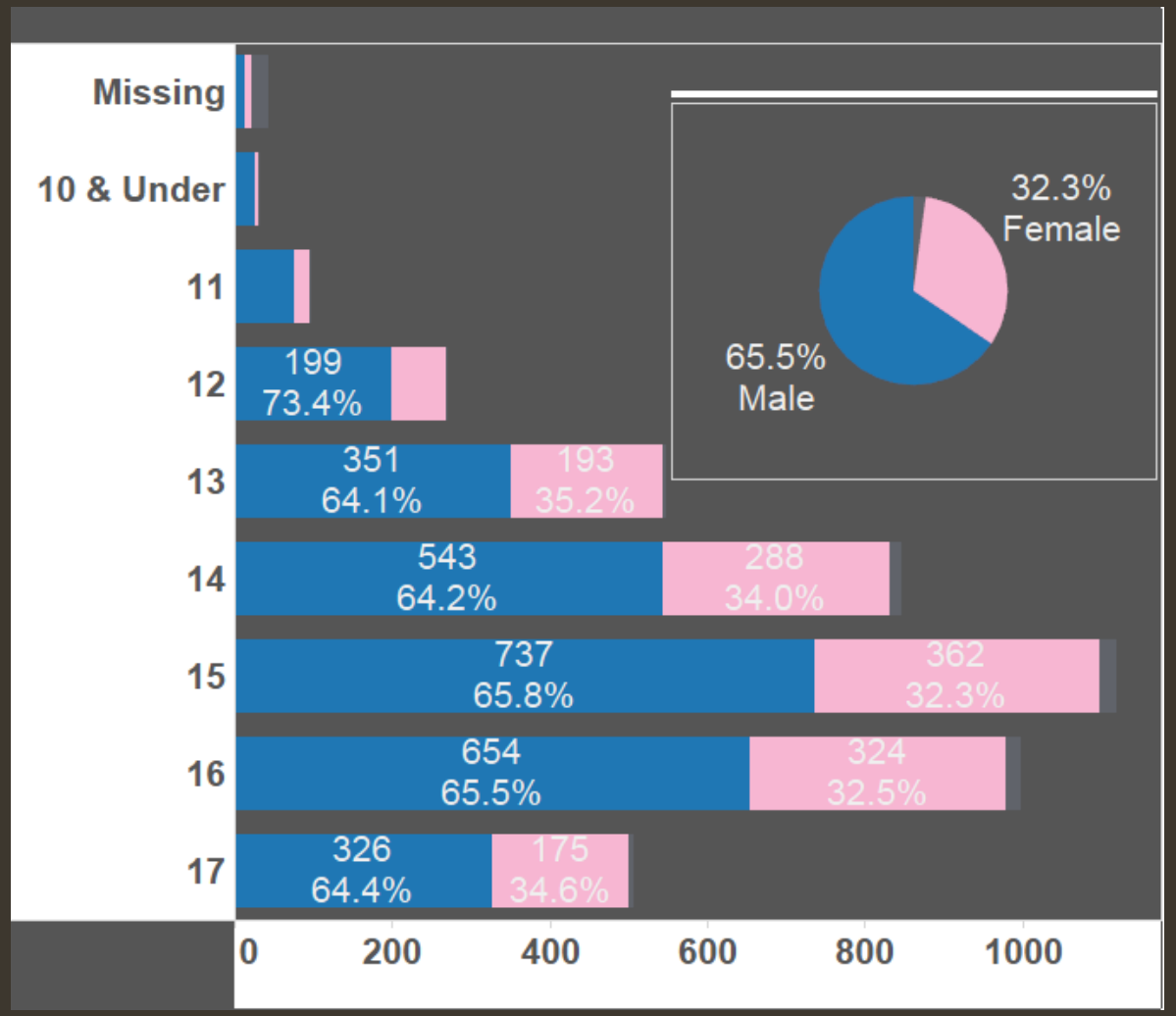
JUSTICE Details

- 1.56 Cases Per Juvenile
- 65.4% Single Case
- 21.1% with 2 cases
- 13.5% with 3 or more cases

Who are they?

- Age at time of First Offense
- Two-third Male
- 1,120 (25.1%) 15 Years of age
- Proportionate Gender Ratio across ages





What did they do?

- 28.5% of the Status Offender Population has a subsequent Misdemeanor or Felony case added later on.
- DMA Taskforce plans on reviewing this in more detail.

  ~ Status to Misd.

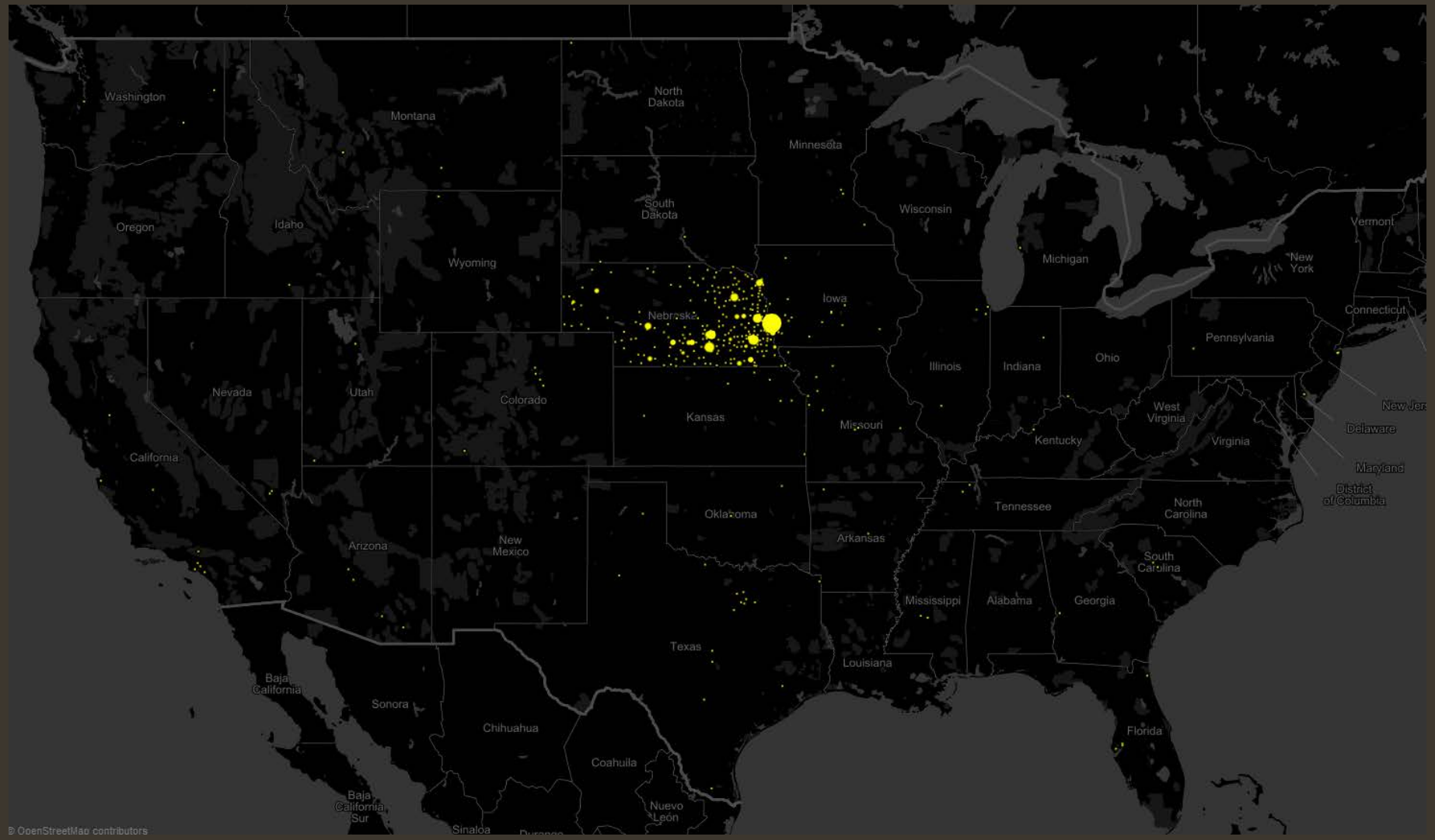
  ~ Misd. to Felony

  ~ etc.…

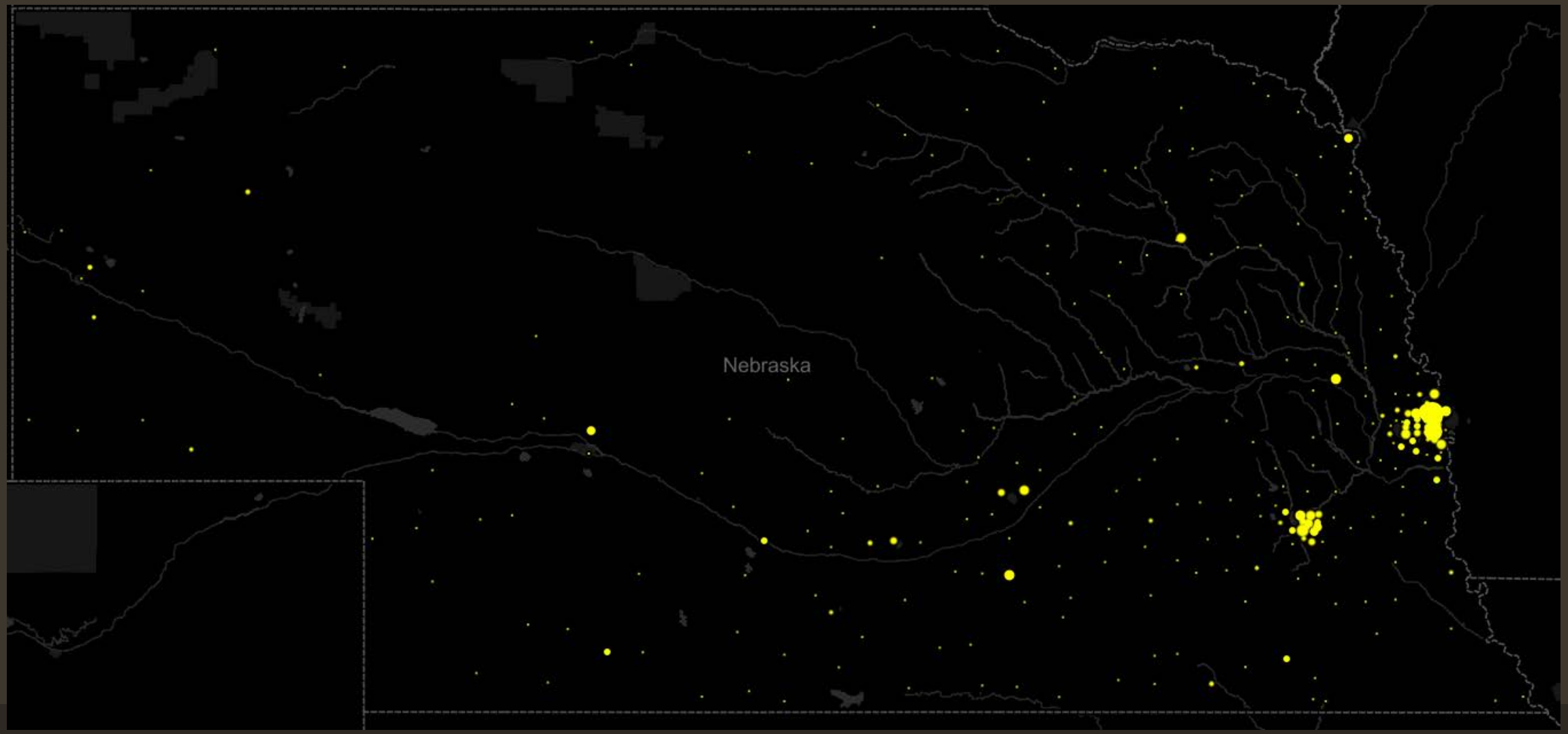
	First Court Sequence	**Most Serious Court Sequence**
Misdemeanor-Infraction	2,383 (53.4%)	2,405 (53.8%)
Status Offender	1,348 (30.2%)	964 (21.6%)
Felony	720 (16.1%)	1,087 (24.4%)
Traffic Offense	13 (0.3%)	8 (0.2%)
Total	4,464	4,464

Where are they from?

- 4,291 from NE (96.1%)
- 125 from Out-of-State (2.8%)
- 48 Missing Address (1.1%)

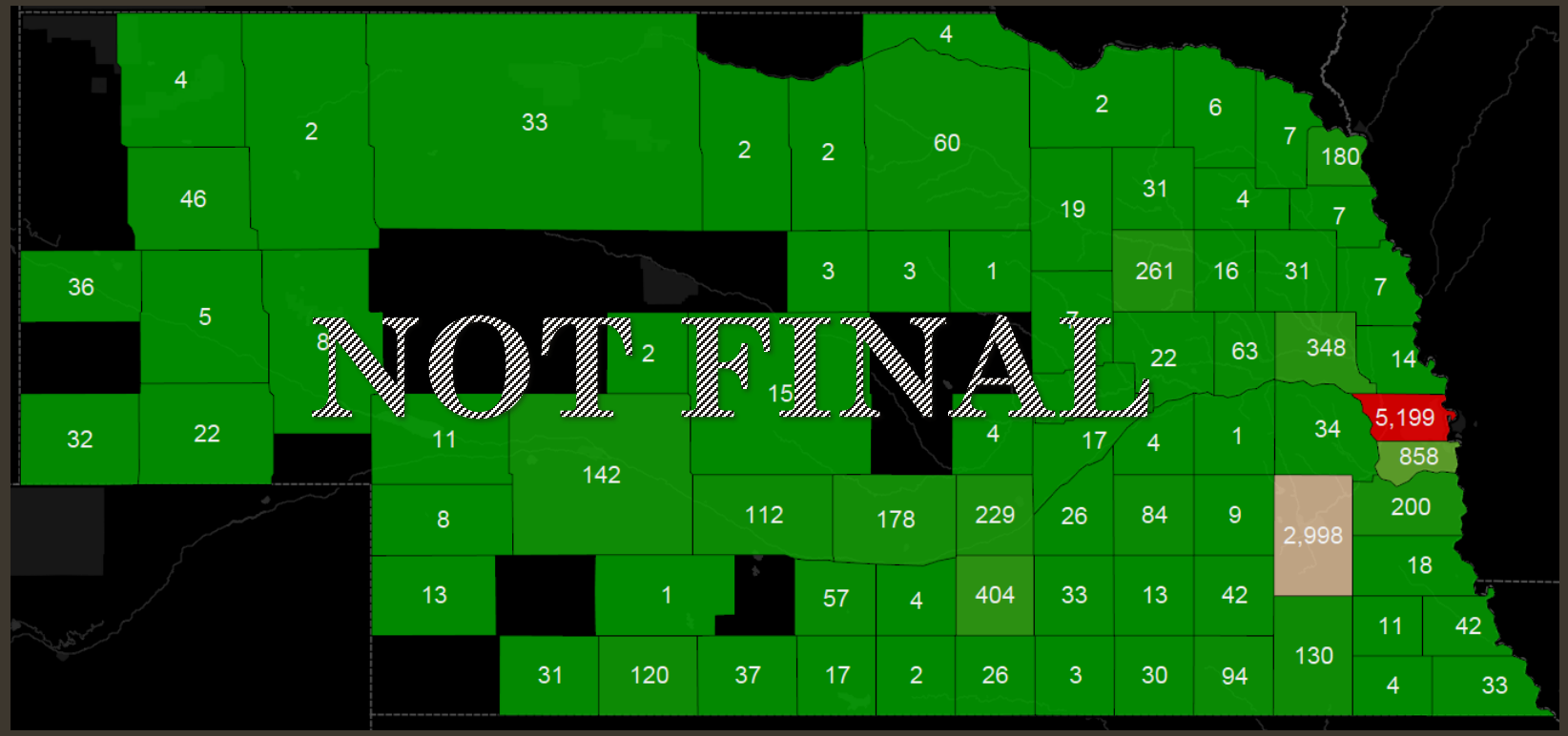


Nebraska up Close



Placement Counts by County (DRAFT)

- Court Cases Breakout
- Douglas 41.3%
- Lancaster 23.8%
- Sarpy 6.8%
- Adams 3.2 %
- Dodge 2.8%
- 22.1% Remaining Counties
- Rates to Follow
- **Difficulty in removing duplicative placements, missing dates, etc.**

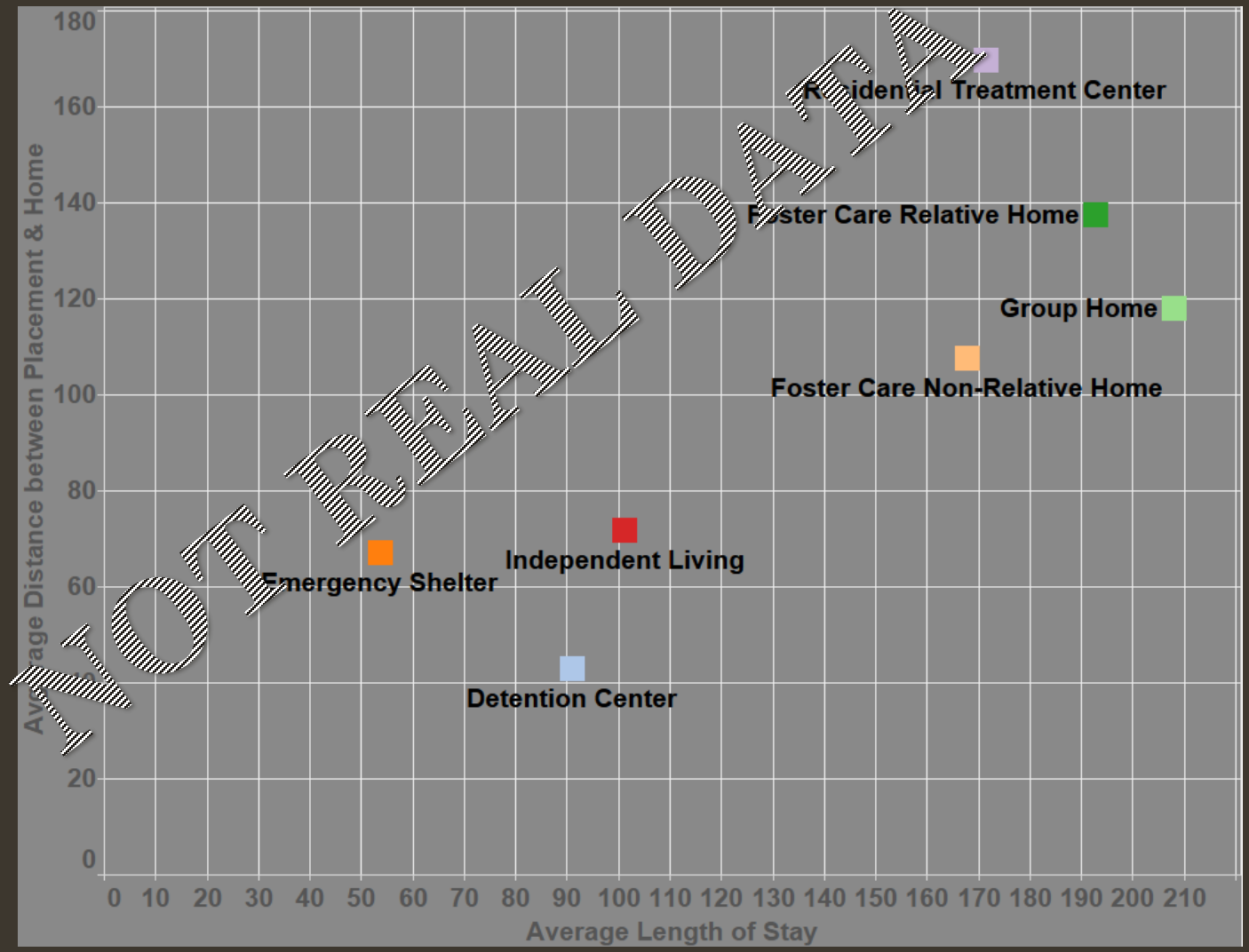


Inconsistency with Data

- Trouble Itemizing Placement Locations
- Re-classify groups
- Grouping Multiple level of Care Facilities
- Tying in additional Data Sources

Look, Think, & Act

- What is next…
- 120 miles for 30 days or 30 miles for 120 days
- Proximity & Duration



Look, Think, & Act

Questions:

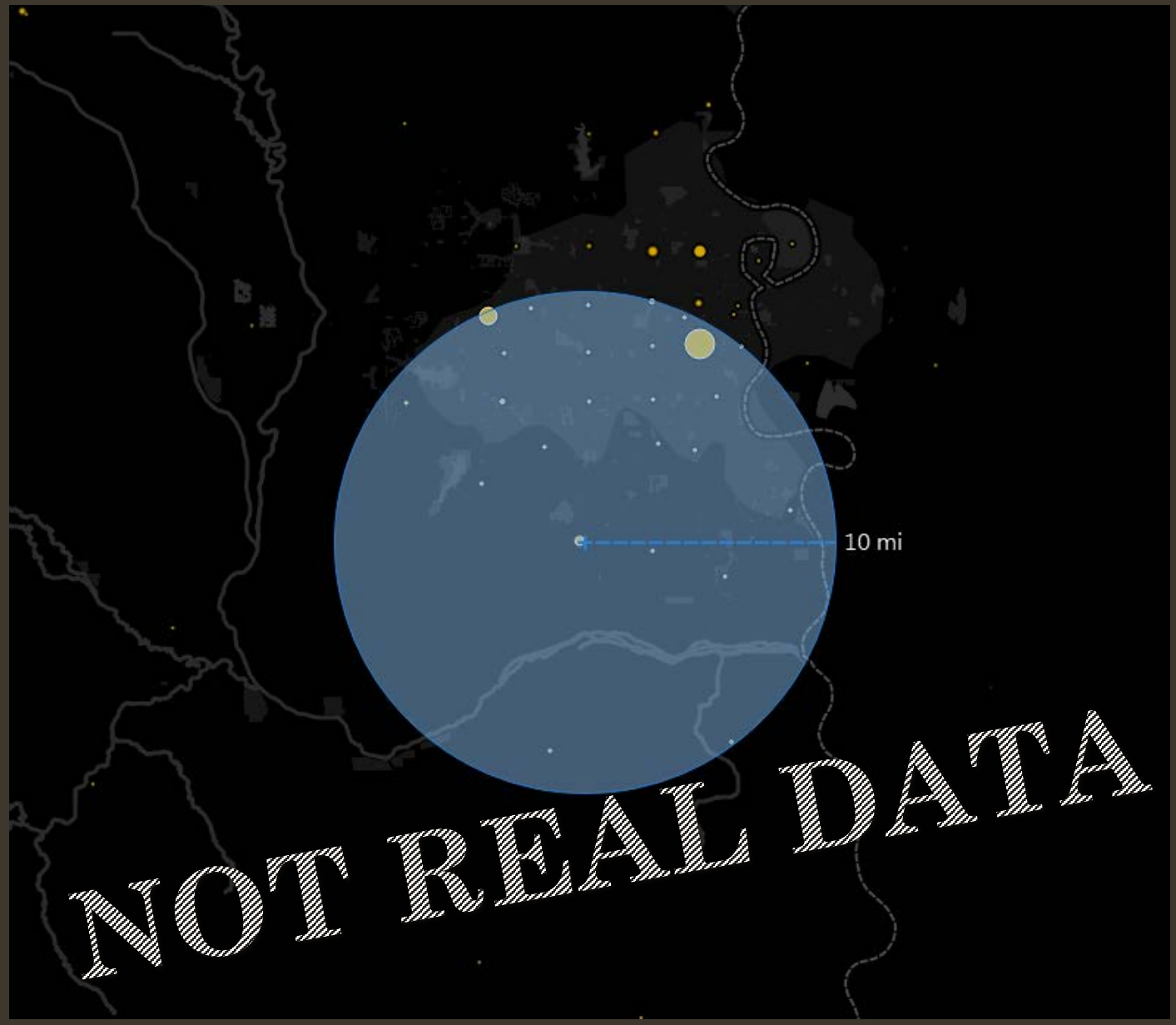
~ Show me all the people within ten miles of _____ that have been in a group home for more than 120 days.

~ Show me how many days have been consumed at the _____ Detention Center, and how far people are having to travel to get there

~ Show me all the placements that…

~ Show me all the cases that…

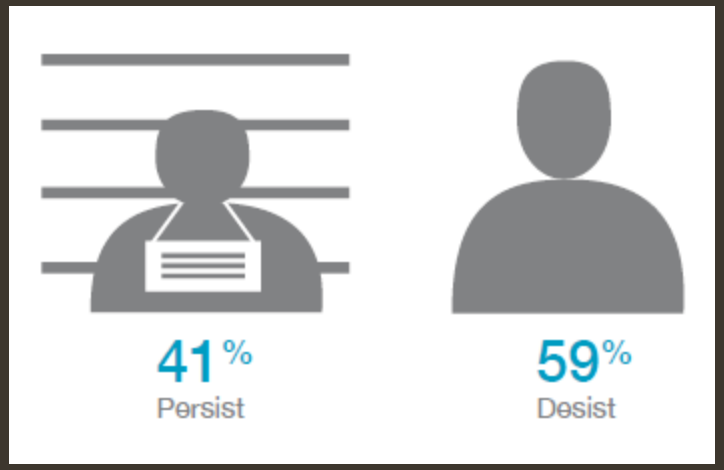
~ Show me all the people that…



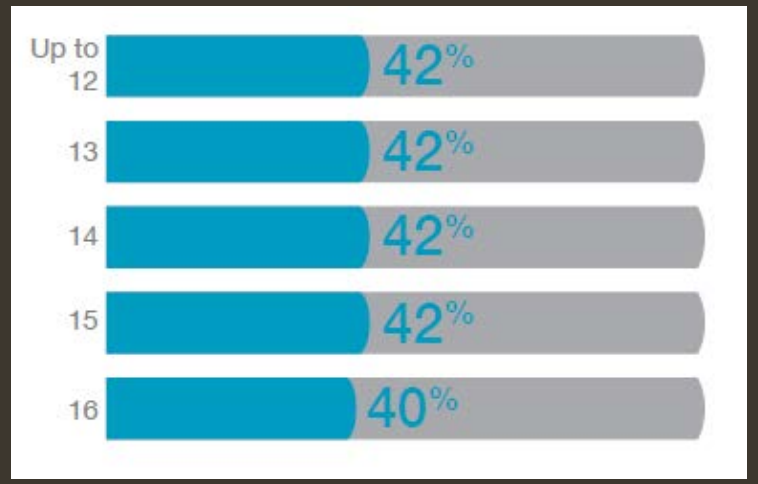
Ideas for Taskforce Goals?

- **Pathways to Desistance (Georgia)**
- **This study describes the likelihood and extent which juvenile offenders persist in illegal behavior and penetrate into the adult system. Linkages were made across multiple agencies to create a longitudinal dataset of hal-million justice-involved individuals spanning five decades.**

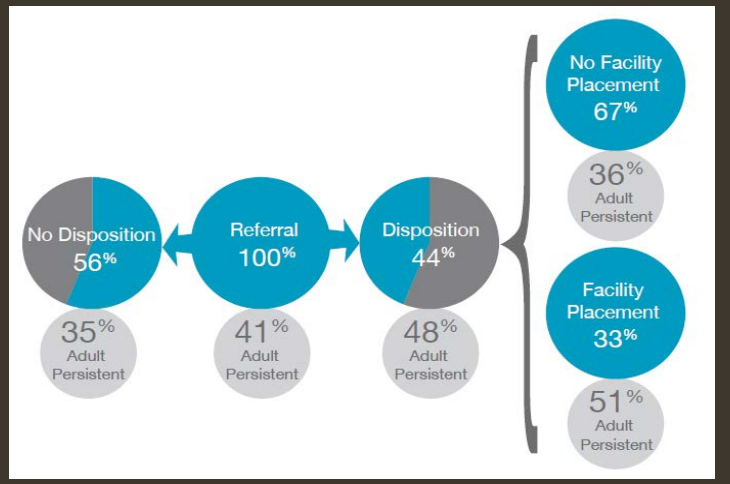
Juveniles That Become Adult persistent



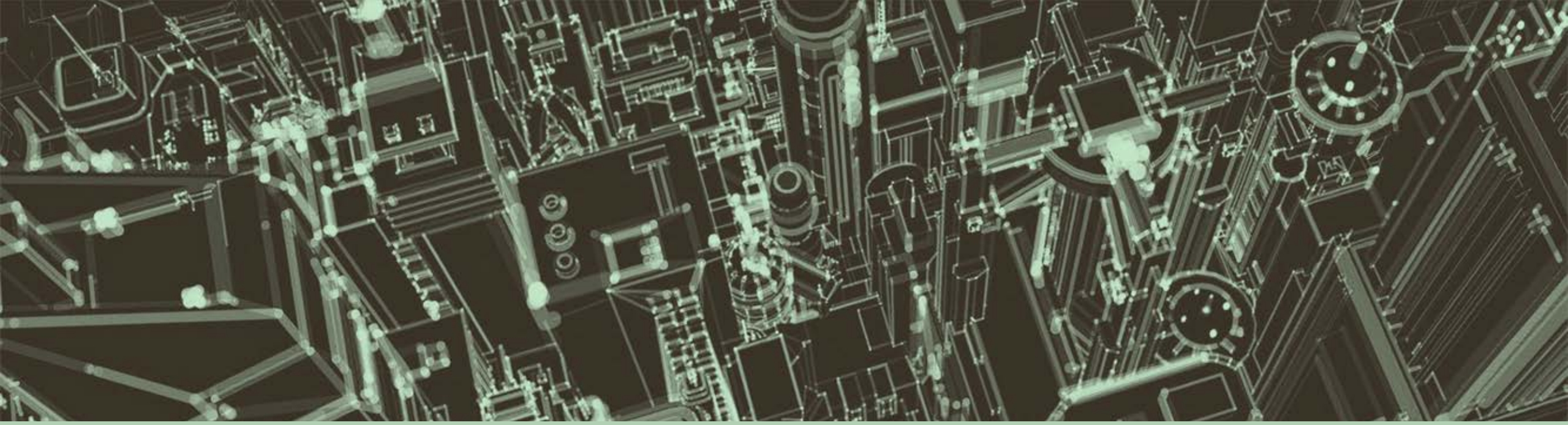
Adult Persistence By Age At First Referral



Adult Persistence at Stage of Intervention



Citation: Pathways to Desistance: Applied Research Services, INC., George Statistical Analysis Center. A Comprehensive Analysis of Juvenile to Adult Criminal Careers. (May 2017) https://cjcc.georgia.gov



Questions