BEYOND HMIS

data science tools ensure equitable & comprehensive reporting of all populations

Elizabeth McDonnell, M.S.
Crossroads Rhode Island

Increasing Capacity &
Building Connections:
Bridging to the Future

the team

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

Elizabeth McDonnell, Data Scientist

Alex Moore, Outcomes & Evaluations

Data work can be lonely; self-supervision is key.

objective

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

Develop a plan to answer, “how many clients were served last year?”

CoC-APR for “Crossroads ALL programs incl HH and Warwick”

5a - Report Validations Table	
Report Validations Table	
1. Total Number of Persons Served	2810
2. Number of Adults (age 18 or over)	2342
3. Number of Children (under age 18)	463
4. Number of Persons with Unknown Age	5
5. Number of Leavers	1956
6. Number of Adult Leavers	1729

This took about 7 minutes to run…

HMIS : when it works & when it doesn't

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

HMIS is great as a service management tool

- ☆ Hosts multi-program data
- ☆ Facilitates coordination of care
- ☆ Organizes HUD-required data
- ☆ Provides some stock reporting

Its not so good as a data analysis tool.

- ✖ Often high financial barrier to entry
- ✖ Steep learning curve
- ✖ Often homeless-client exclusive
- ✖ Hard to customize (proprietary)
- ✖ Aggregate reporting VERY SLOW

outline

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

- Rhode Island & Crossroads programs
- Populations missing from HMIS-centered analyses
- Experience with database merges.
- Insights from multiple data sources.

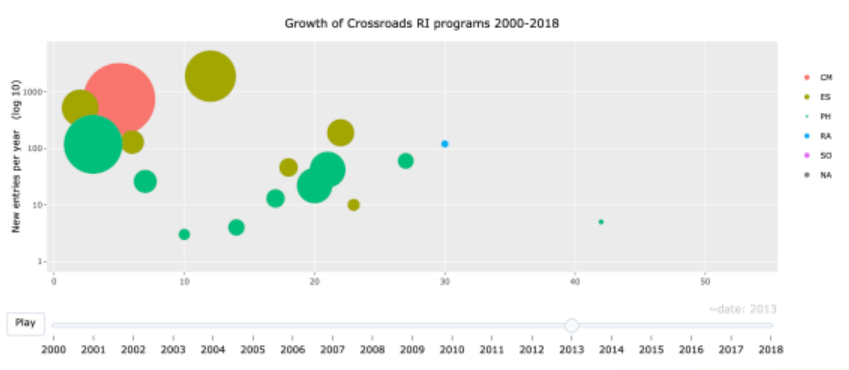
CoCs	Total homeless	Unsheltered Homeless	Homeless people In families	Sheltered people in families
1	1, 180	69	378	378

Crossroads provides ~ 80% of homeless services within the RI CoC

programs: housing first

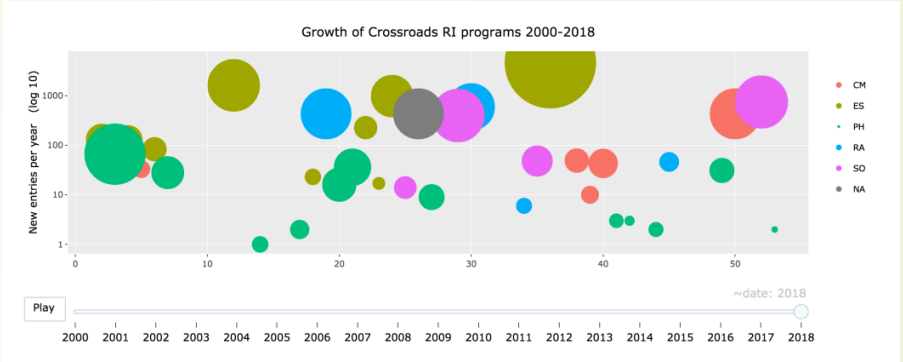
Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

2013



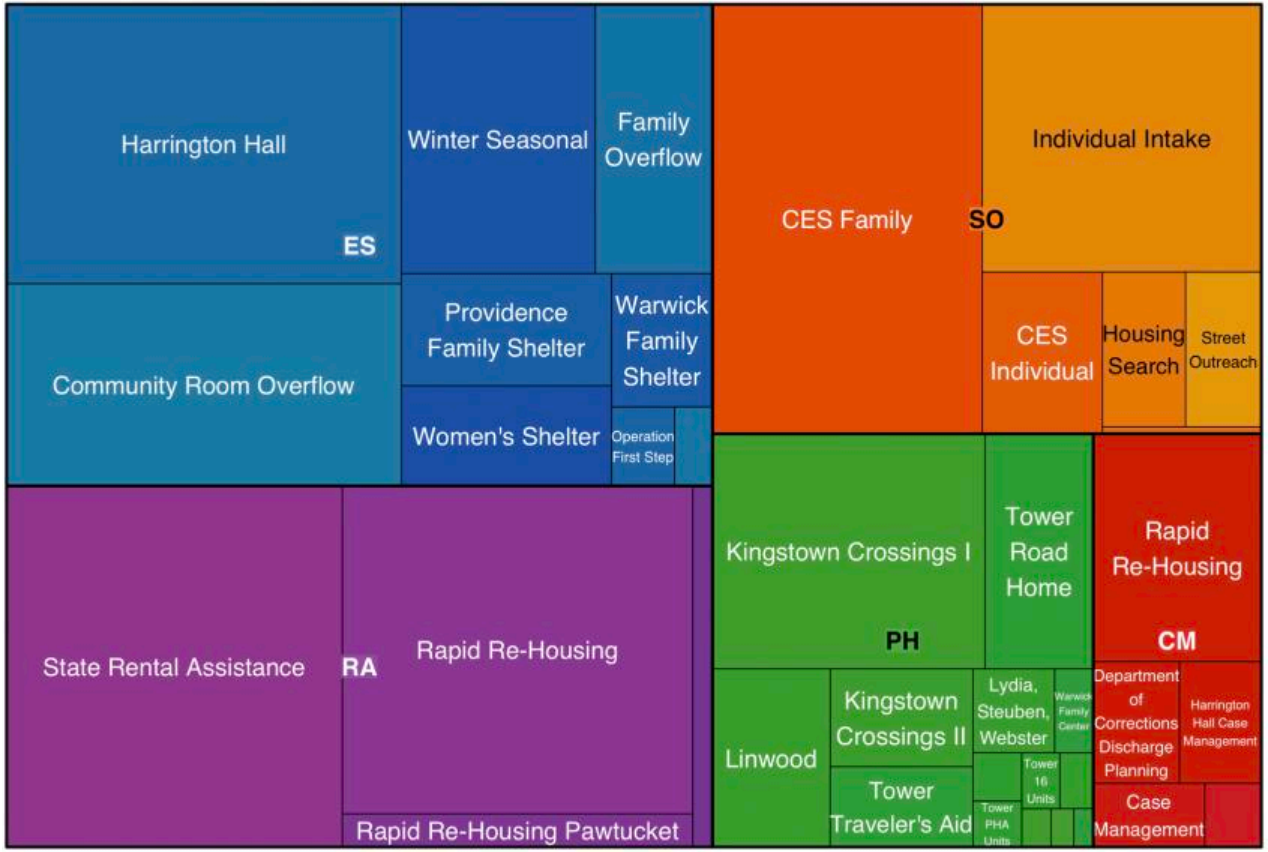
Shelter & permanent housing focused

2018



Lots of case management & services

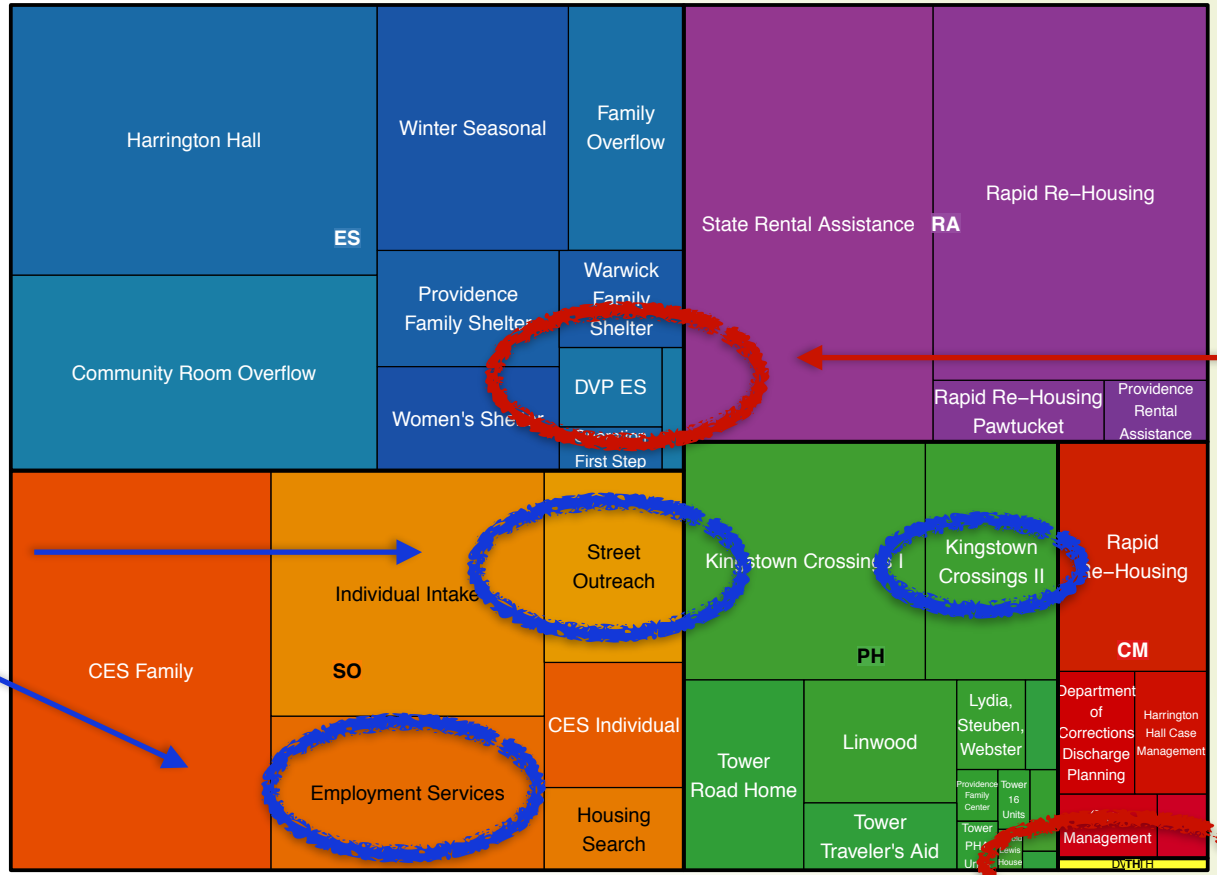
3,988
de-duped clients across programs.



Crossroads Rhode Island Program Overview

4,596
de-duped clients across programs.

Clients not in HMIS are accounted for.



Domestic violence programs are represented.

Crossroads' databases

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

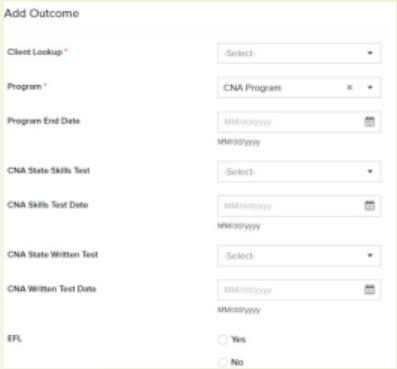
EmpowerDB

mri REAL ESTATE SOFTWARE Bostonpost

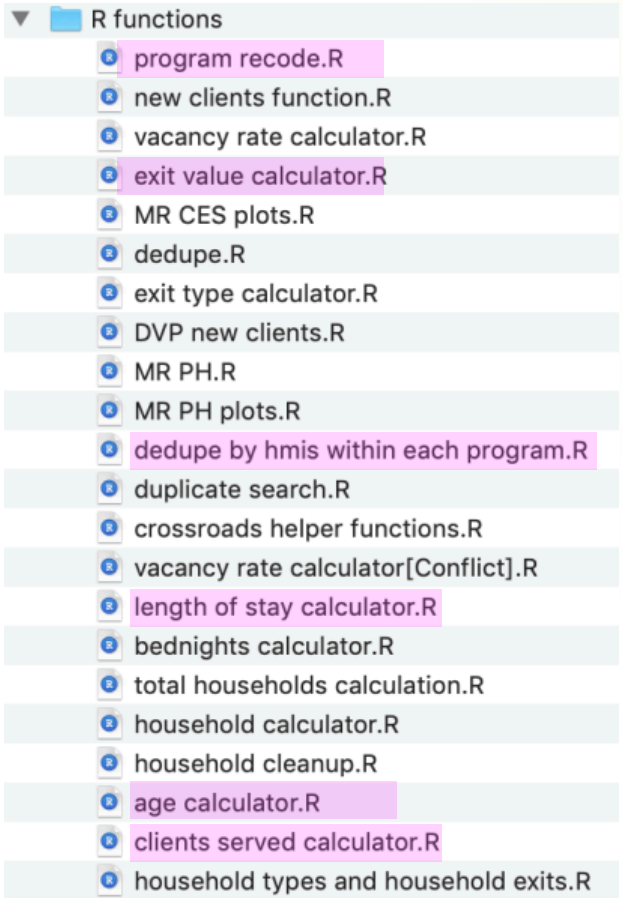
HMIS

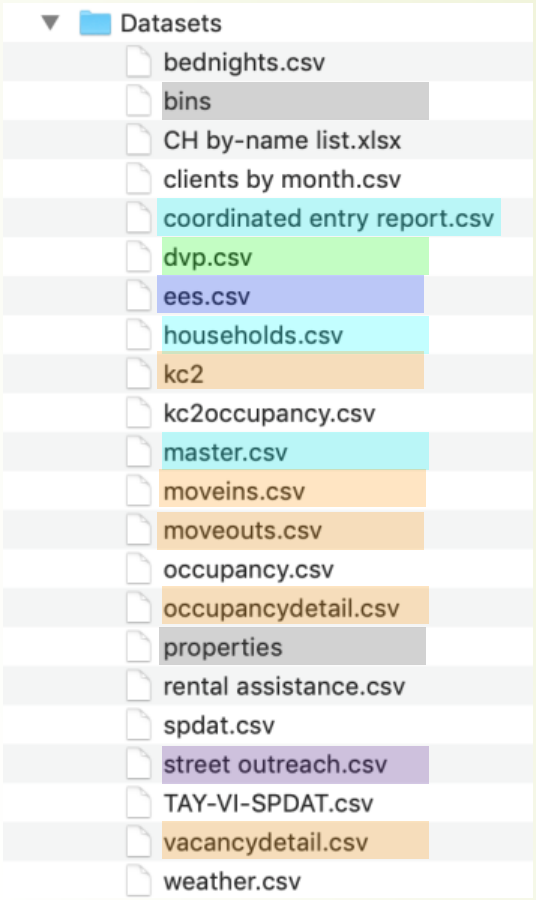
street outreach

Date	Location	Downtown?	Last Name	First name	HMIS ID #	Type of Contact
7/2/18	Trinity Square (Prov.)	No	[redacted]		0	Client Refused Co
7/2/18	Weybosset Street (Prov.	Yes	[redacted]		0	Actual Contact (m
7/2/18	Kennedy Plaza	Yes	[redacted]		0	Actual Contact (m
7/5/18	Kennedy Plaza	Yes	[redacted]		47567	Actual Contact (m
7/6/18	Weybosset Street (Prov.	Yes	[redacted]		34619	Actual Contact (m
7/6/18	Waterplace Park (Prov.)	Yes	[redacted]		13125	Actual Contact (m



EES





- HMIS report
- Created by me
- BostonPost report
- Street outreach excel file
- EmpowerDB custom report
- EES excel file

a favorite: determining clients served

```
## identifies all clients served during a specified time period
## conditional dedupe option applies to data with either
# an hmis_id or a name_id

clients_served <- function(df, end_date, start_date, dedupe=TRUE){
  newdf <- data.frame(df[df$entry_date <= end_date &
                           (is.na(df$exit_date) | (df$exit_date >= start_date)), ])

  if(dedupe){
    if("hmis_id" %in% colnames(df)){
      uniquedf <<- newdf %>%
        group_by(hmis_id) %>%
        slice(which.max(as.Date(entry_date, "%m/%d/%Y")))
    }
    else {uniquedf <<- newdf[!duplicated(newdf$name_id), ]}
  }

  else {newdf <<- newdf}

}
```

steps for processing data

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

- Begin with master HMIS custom report - all entry/exits with no date limits.
- RECODE most variables so they make sense to regular people:

programs
```
levels(data$bin)[levels(data$bin)=="Crossroads - Rhode Island Family Shelter (ES-FAM) (CHF)"] <- "1346"
```
race
```
levels(clients$primary_race)[levels(clients$primary_race)=="Client refused (HUD)"] <- "Other"
levels(clients$primary_race)[levels(clients$primary_race)=="White (HUD)"] <- "White"
```
veteran
```
levels(clients$veteran)[levels(clients$veteran) =="No (HUD)"] <- "No"
```

- Merge with separate "bins" dataframe to add program info (i.e., ES or PH, etc.)

```
clients <- merge(clients, bins, by = "bin")
```

- Add calculated fields to aid in analysis: age at entry, length of stay, exit value, is child, status (ope
- Create a variable — key — that can be used to merge with other databases (name_id)

```
clients$name_id <- paste(clients$first_name, clients$last_name, clients$birth_date)
```

bin	hmis_id	household_id	last_name	first_name	entry_date	exit_date	birth_date	gender	primary_race
1332	38865	6769	[redacted]		2012-01-12	2012-03-14	[redacted]	Male	Black or African American
1332	21877	3665			2007-08-28	2007-10-12		Male	White

...

program	support_type	people_served	active	capacity	age	entry_age	status	los	is_child	exit_type	exit_value	name_id
Providence Family Shelter	ES	FAM	YES	45	33	26	Exited	62	No	Emergency shelter	negative	[redacted]
Providence Family Shelter	ES	FAM	YES	45	41	29	Exited	45	No	Emergency shelter	negative	

84,346 entry/exits to HMIS programs

clean remaining datasets

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

EES

- Split full name into first_name and last_name
- Recode ‘m’, ‘f’, ‘w’ to Male or Female
- Recode ‘af’, ‘aa’, ‘bl’ to Black or African American
- Remove rows with notes-to-self

KC2

- Split full name and remove commas, initials etc
- Anti_join BostonPost & HMIS clients to get KC2 clients not in HMIS.
- Predict gender & race from first and last names

DVP

- Convert # kids from string (“3 kids”) to integer
- Remove rows where kids are counted twice (long story)
- Assign program based on room number on file

CES

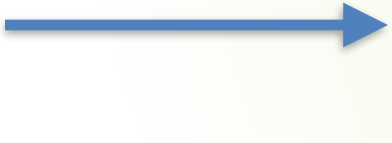
- Write & use lookback function to find clients new in the current month with prior calls
- Use NLP on case notes for no contact > 90 days
- Identify shelter exits from 3 distinct fields

Kingstown Crossings II

How many clients were served last year?

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

MRI AFFORDABLE HOUSING BOSTONPOST



HMIS clients

- housed anytime in 2018
- rich hmis-based dataset

NOT in HMIS

- in housing on last day of 2018
- thin dataset
- requires more steps

EES
- Create name_id

KC2
- Create name_id

DVP
- Create name_id

CES
- All have hmis_id already & name_id

- JOIN client data from above to master client data using name_id
- Dedupe as needed.

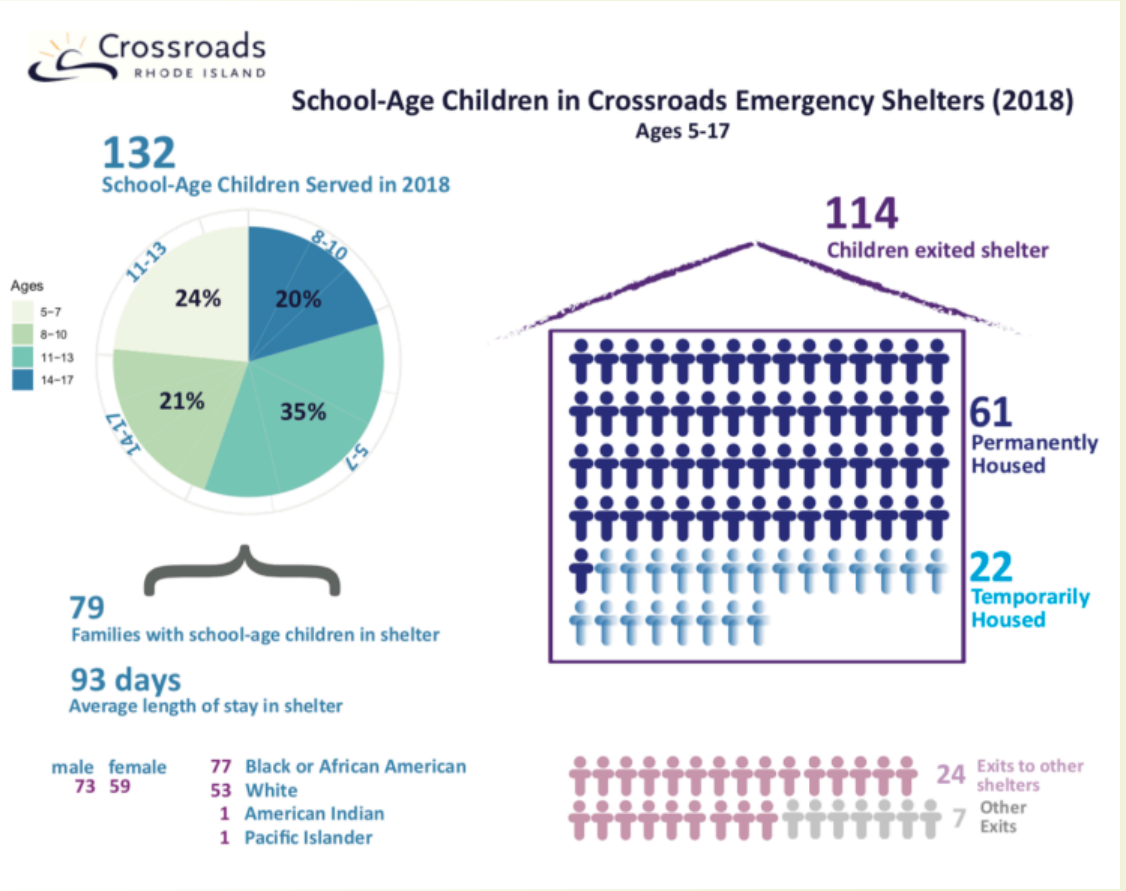
tips for data cleaning

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

- Write independent functions that can apply across datasets.
  - The more functions, the better.
- STANDARDIZE ALL VARIABLE/COLUMN NAMES
- Keep like-data in separate tables. Take advantages of joins.
  - i.e. bin df separate from entry/exit df
- Supervise your own work - double check everything by hand/ eye.
- Be able to trace each data point (client, entry, etc.) through dataframes. Identify when they're duplicated and de-duplicated. Know where your data is and is not.

infographic requires:

- Cleaned & recoded master clients dataframe
- Calculated fields : age range, exit types, length of stay.
- Households dataframe

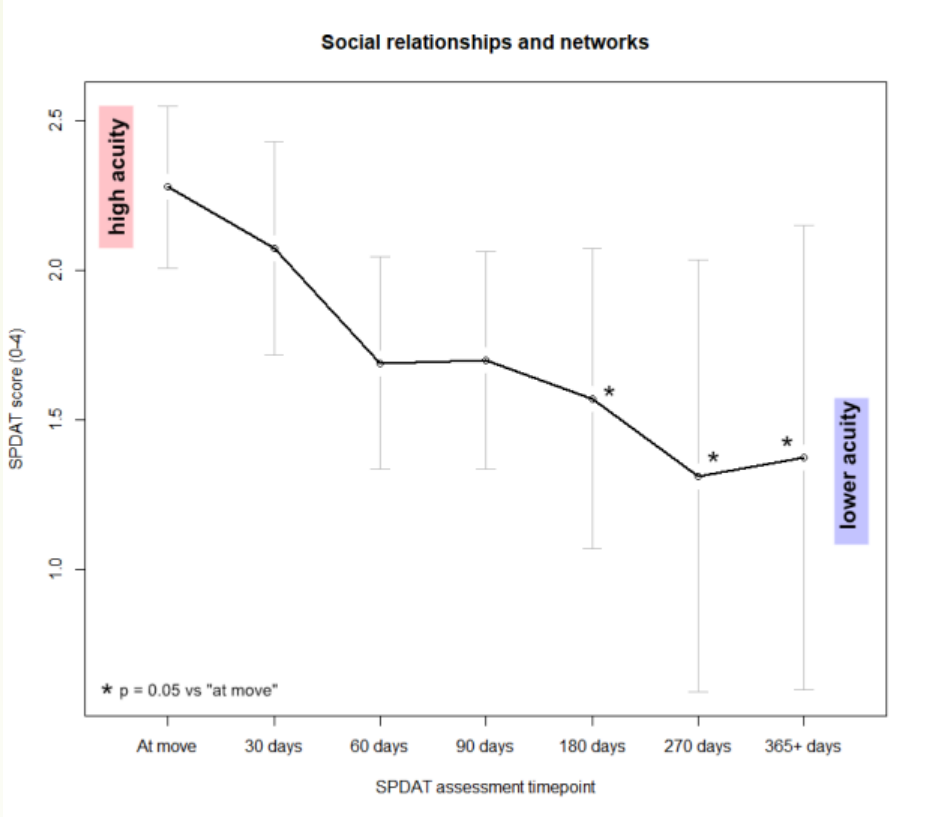


plot requires:

- [illegible]nts dataframe
- SPDAT score custom report
- bins / program dataframe

then:

- Filter out most recent test date, anti join with df, continue filtering until all tests are ordered then bind back together
- Normalize to housing move-in date & run significance testing



funding is given to data-driven organizations

- Many small homeless service-based programs do not use HMIS.
- Many organizations that DO use HMIS also serve clients NOT in HMIS.

- Data science tools are flexible, often free or low-cost, and very FAST compared to HMIS reporting.
- Beyond required standards, third-party software is not only required for accurate internal reporting, but allows exploration and insights that would be impossible using HMIS alone.

questions?

Elizabeth McDonnell, M.S.
Crossroads Rhode Island
emcdonnell@crossroadsri.org

BEYOND HMIS

data science tools ensure equitable & comprehensive reporting of all populations

Elizabeth McDonnell, M.S.
Crossroads Rhode Island

Increasing Capacity &
Building Connections:
Bridging to the Future

the team

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

Elizabeth McDonnell, Data Scientist

Alex Moore, Outcomes & Evaluations

Data work can be lonely; self-supervision is key.

objective

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

Develop a plan to answer, “how many clients were served last year?”

5a - Report Validations Table	
Report Validations Table	
1. Total Number of Persons Served	**2810**
2. Number of Adults (age 18 or over)	**2342**
3. Number of Children (under age 18)	**463**
4. Number of Persons with Unknown Age	**5**
5. Number of Leavers	**1956**
6. Number of Adult Leavers	**1729**

CoC-APR for “Crossroads ALL programs incl HH and Warwick”

HMIS : when it works & when it doesn't

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

PROS

- Hosts multi-program data
- Facilitates coordination of care
- Organizes HUD-required data
- Provides some stock reporting

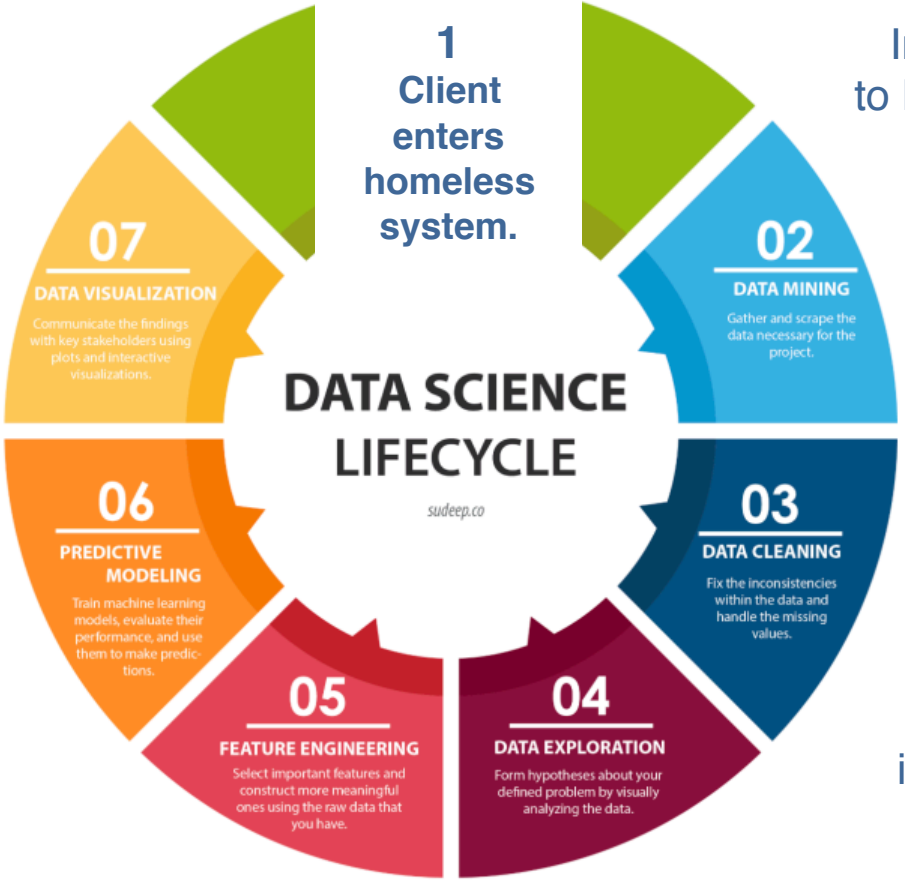
CONS

- High financial barrier to entry
- Steep learning curve
- Homeless-client exclusive
- Difficult to anonymize
- Hard to customize (proprietary)
- Aggregate reporting VERY SLOW

outline

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

- What is data science, and why bother?
- Rhode Island, Crossroads, and housing first
- HMIS : when it works and when it doesn't
- Populations missing from HMIS-centered analyses
- Basics of database merges
- How many clients did Crossroads serve last year?



Initial data is entered to HMIS and iteratively updated.

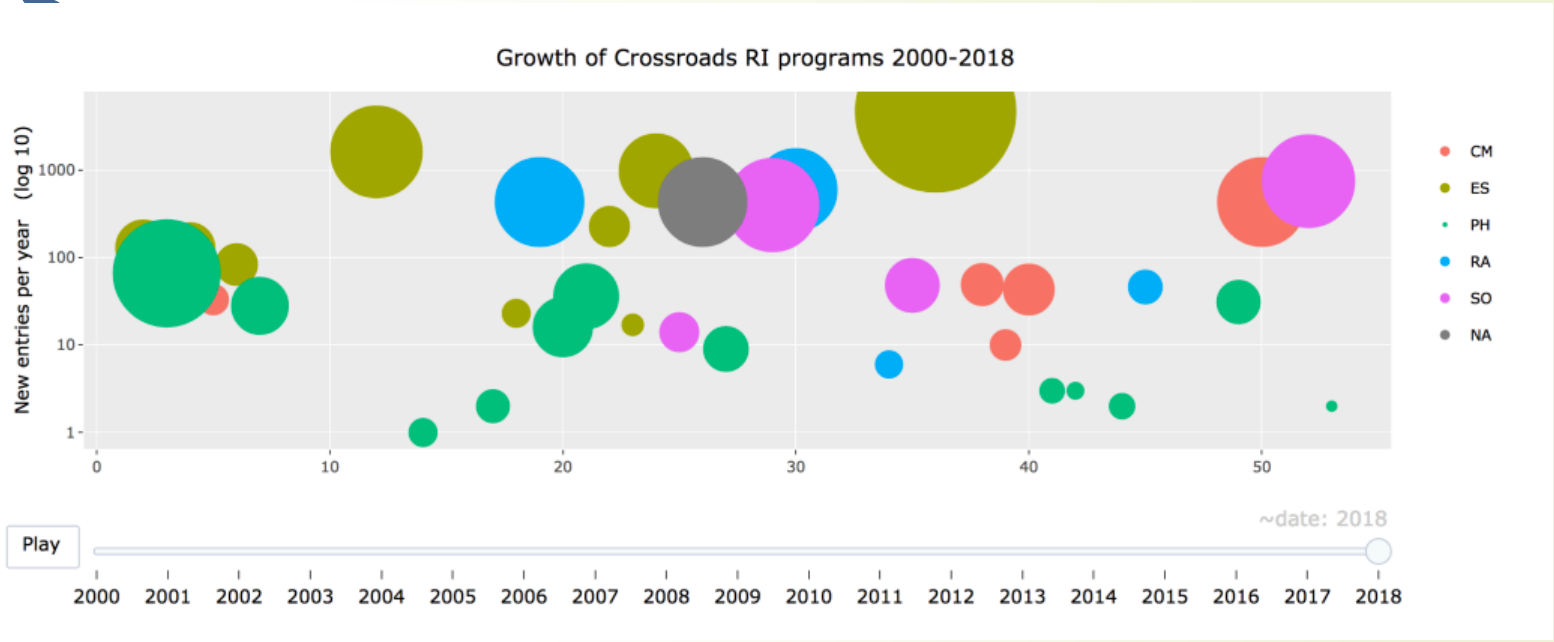
Data team pulls data into analysis software to drive insights.

CoCs	Total homeless	Unsheltered Homeless	Homeless people In families	Sheltered people in families
1	1, 180	69	378	378

Crossroads provides ~ 80% of homeless services within the RI CoC

program
s

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS



programs

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

shelter

Operation first step
Women's shelter
Harrington hall
Community room overflow
Winter seasonal
Family overflow
Couple's shelter
Warwick family shelter
Providence family shelter
Domestic violence shelter

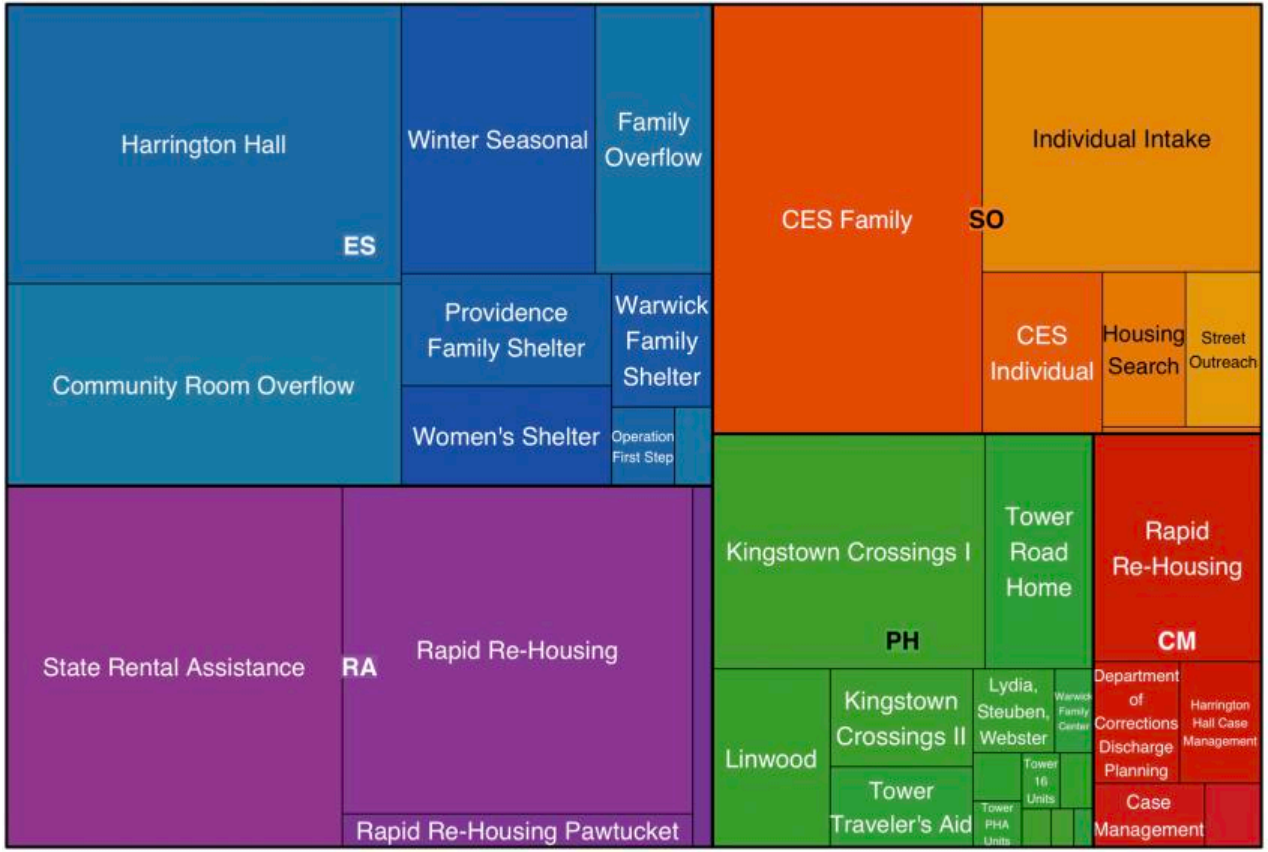
permanent housing

Kingstown crossings I
Kingstown crossings II
Harold Lewis House
Tower units
Providence Family Center
Warwick Family Center
Mike Terry Apartments
Crossroads Family Housing
Tremont St. Apartments

other supportive programs

Employment & education services
Rental assistance
Street outreach
Coordinated entry

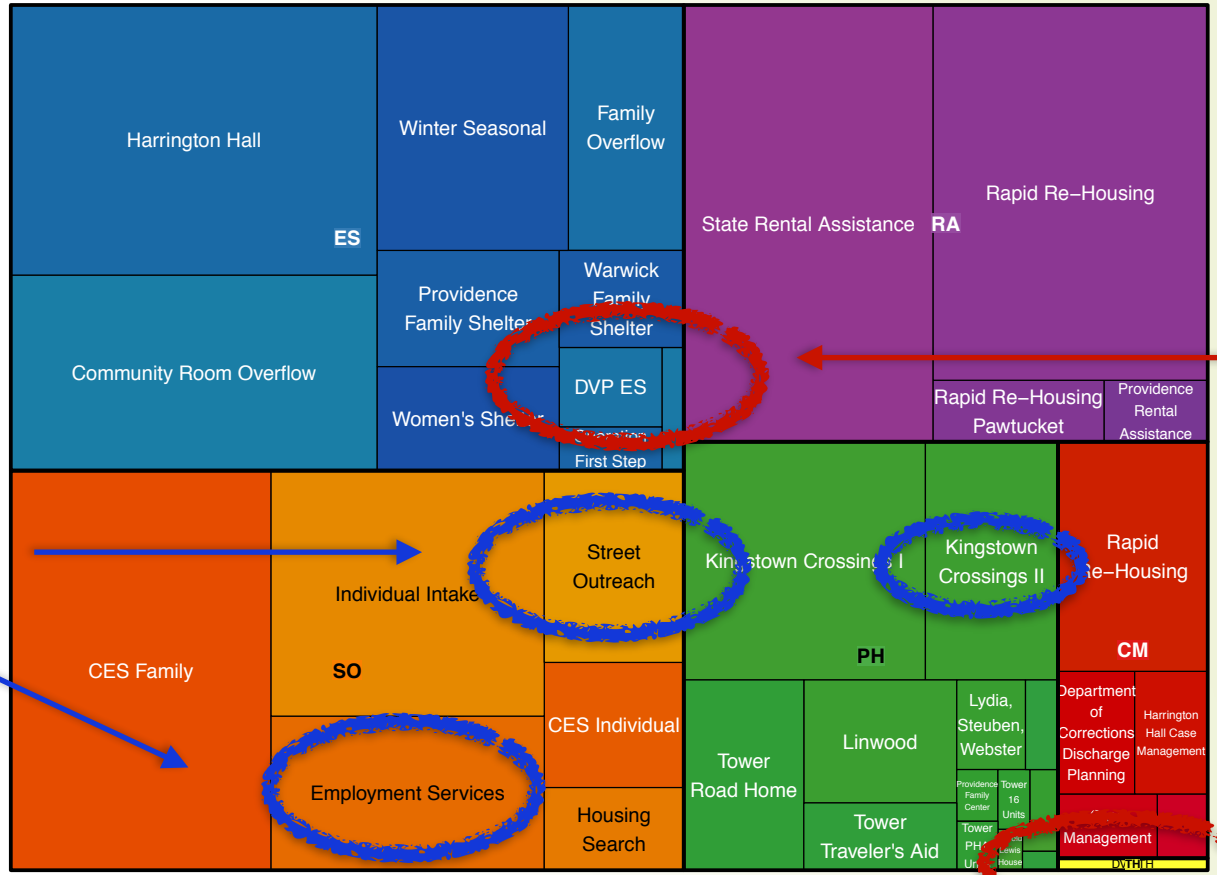
3,988
de-duped clients across programs.



Crossroads Rhode Island Program Overview

4,596
de-duped clients across programs.

Clients not in HMIS are accounted for.



Domestic violence programs are represented.

excluded populations

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

The data we were missing was not independently distributed across populations.

- All sheltered female survivors of domestic violence & their children.
- Approximately half of the families served in one of our PH locations
- Many street-level contacts with homeless individuals
- Almost all clients who sought employment & education services.

If we fail to report on these populations, (1) internal reporting will be inaccurate and (2) program-level decisions will not be data-driven.

Crossroads' databases

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

EmpowerDB

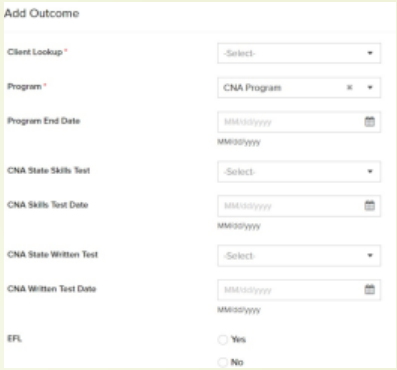
mri REAL ESTATE SOFTWARE Bostonpost

SERVICEpoint®

street outreach

Date	Location	Downtown?	Last Name	First name	HMIS ID #	Type of Contact
7/2/18	Trinity Square (Prov.)	No	[redacted]		0	Client Refused Co
7/2/18	Weybosset Street (Prov.	Yes	[redacted]		0	Actual Contact (m
7/2/18	Kennedy Plaza	Yes	[redacted]		0	Actual Contact (m
7/5/18	Kennedy Plaza	Yes	[redacted]		47567	Actual Contact (m
7/6/18	Weybosset Street (Prov.	Yes	[redacted]		34619	Actual Contact (m
7/6/18	Waterplace Park (Prov.)	Yes	[redacted]		13125	Actual Contact (m

EES



basics of merging

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

- Related data is stored in multiple tables and referenced by common "keys."

UDEs

program entries

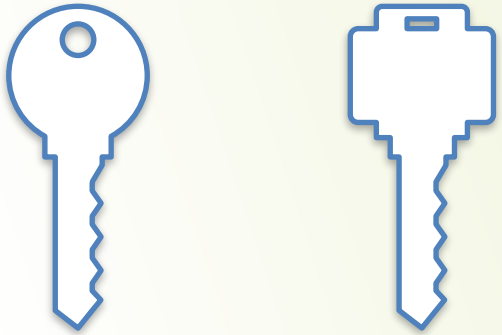
	hmis_id	household_id	last_name	first_name	birth_date	gender	primary_race
137	61998	9697	[redacted]	[redacted]	[redacted]	Male	White
221	61999	9697	[redacted]	[redacted]	[redacted]	Male	White
629	60218	9494	[redacted]	[redacted]	[redacted]	Female	Black or African American
821	[redacted]	6671	[redacted]	[redacted]	[redacted]	Female	Black or African American
836	7281	1549	[redacted]	[redacted]	[redacted]	Female	White

Each HMIS id is related to one or more program entry/exit records.

bin	hmis_id	entry_date	exit_date	exit_destination	program
1867	54470	2017-10-27	2017-11-06	Staying or living with friends, temporary tenure (e.g., ...	State Rental Assistance
1869	54470	2017-08-29	2017-11-06	No exit interview completed (HUD)	Rapid Re-Housing
1891	54470	2019-01-15	2019-01-15	Staying or living with friends, temporary tenure (e.g., ...	CES Family
1332	54471	2019-03-25	*NA*	*NA*	Providence Family Shelter
1891	54471	2019-02-08	*NA*	*NA*	CES Family
1332	54471	2016-07-26	2016-11-02	Transitional housing for homeless persons (including ...	Providence Family Shelter
1548	54471	2016-07-13	2016-07-26	Emergency shelter, including hotel or motel paid for ...	Family Overflow
1819	54471	2017-10-03	2017-10-27	Other (HUD)	State Rental Assistance

Crossroads RHODE ISLAND
Helping those in need for 125 YEARS

What if the keys don't match?



EmpowerDB

Domestic Violence Program : Unique Clients Served	2018 ES	2018 TH	2018 Total
Total unique clients served, n (%)	**92**	**35**	**127**
Adult females	45 (49)	15 (43)	60 (47)
Adult transgender (MTF)	1 (< 1)		1 (< 1)
Children	46 (50)	20 (57)	66 (52)
Age range (all unique adult clients), n (%)	**46**	**15**	**61**
18-24	8 (17)	2 (13)	10 (16)
25-34	18 (39)	7 (47)	25 (41)
35-44	14 (31)	3 (20)	17 (28)
45-54	4 (9)	2 (13)	6 (10)
55-61	2 (4)	1 (7)	3 (30)
Primary race (all unique adult clients), n (%)	**46**	**15**	**61**
American Indian or Alaska Native	1 (2)	1 (7)	2 (3)
Black or African American	13 (28)	5 (33)	18 (30)
Native Hawaiian or Other Pacific Islander	1 (2)		1 (2)
White	22 (48)	5 (33)	27 (44)
Other	9 (20)	4 (27)	13 (21)
Total households, n (%)	**46**	**15**	**61**
Unaccompanied adults with no children	20 (43)	6 (40)	26 (43)
Single parent with children	26 (57)	9 (60)	35 (57)

Some of these clients are also in HMIS. We don't want to count them twice!

Create one!

Find relationship between datasets. What's in common?

This is only necessary if there are clients represented in both datasets that need to be de-duplicated.

Solution: name_id

```
age <- function(data){⟺}
length_of_stay <- function(data){⟺}
exit_type <- function(data) {⟺}

dvp <- read_csv("C:/Users/emcdonnell/Google Drive/CrossroadsRI/General Data/Datasets/dvp.csv")
bins <- read_csv("C:/Users/emcdonnell/Google Drive/CrossroadsRI/General Data/Datasets/bins.csv")
DVP <- dvp

#change column names
DVP <- setNames(DVP, c("name","is_child","children", "entry_date", "exit_date", "program",
                       "exit_destination", "exit_reason", "birth_date", "primary_race", "gender",
                       "homeless", "veteran", "disability", "previous_residence"))
```

- clean the data
- clean the data
- clean the data
- clean the data

```
DVP <- extract(DVP, name, c("first_name", "last_name"), "([^ ]+) (.*)")
DVP$bin <- DVP$program
DVP$name_id <- paste(DVP$first_name, DVP$last_name, DVP$birth_date)
```

duplicate check: DVP

```
DVP_2018 <- clients_served(DVP, "2018-12-31", "2018-01-01", dedupe = FALSE)
dedupe_program_name_id(DVP_2018)
```

```
all_2018 <- rbind.fill(clients_2018, DVP)
```

```
all_2018 <- all_2018[!duplicated(all_2018$name_id), ]
```

Crossroads Emergency Shelter Programs: 2018 Summary Data

Capacity: 327 individual & family units

	Program	Capacity (units)	Total clients served	Bednights
	Emergency Shelter programs	**327**	**2,814***	**106,093**
IND	Operation First Step	20	41	3,625
IND	Women's Shelter	41	169	12,174
IND	Harrington Hall	112	884	38,748
IND	Community Room Overflow	16	640	9,735
IND	Winter Seasonal	85	419	4,365
FAM	Family Overflow	13	249	2,726
FAM	Couple's Shelter	4	23	2,750
FAM	Warwick Family Shelter	9	108	9,477
FAM	Providence Family Shelter	15	189	18,111
FAM	Domestic Violence Program (ES)	12	92	4,382

* Clients de-duplicated by program but not across shelter programs. See data below for unique clients across shelter programs.

ES programs : Unique Clients Served	2018
Total unique clients served, n (%)	**2,038**
Adult males	1,228 (60)
Adult females	542 (27)
Adult transgender	11 (< 1)
Male children	135 (7)
Female children	122 (6)

Table 1 Clients served across all programs	2018 * does not include coordinated entry
Total unique clients served, n	**3,549**
Adult males	1,688
Adult females	981
Adult transgender/non-conforming	16
Adult unknown/data not collected	134
Male children	381
Female children	349

- Many organizations have “secondary” databases where a much smaller amount of client data is collected and stored.
- This data is hard to incorporate into organization-wide reporting and requires a few additional programming steps. This makes it easy to skip including this data in broader reporting.
- Data in secondary databases represents special populations — this is why they’re in a secondary database! Whether DV victims, homeless sheltered at a small parochial shelter, or those receiving additional support services, skipping over reporting on these populations excludes them from data-driven insights.

questions?

Elizabeth McDonnell, M.S.
Crossroads Rhode Island
emcdonnell@crossroadsri.org