Hudson River Housing 2023 HMIS Data

Vassar College Math-301

Alesio Dosti, Audrey Myers, and Ian Zumpano

April 2024

Contents

Introduction

In this report, we explore 2023 data from the Homeless Management Information System (HMIS) of Hudson River Housing. This data includes any individual who was in the system during the year of 2023. It is important to note that this dataset only included one entry per client ID. This means that if a resident were to be admitted and discharged multiple times within the year, this dataset would only include one of those admissions/discharges.

We visualize the demographics of guests and residents served by Hudson River Housing, investigate length of program stay across programs, focusing on emergency, transitional, and permanent supportive housing, and explore the data quality.

Demographic Information

The greatest proportion of individuals in Hudson River Housing's system in 2023 were in an emergency housing program (Figure 1). Figures 2-6 show that the population that HRH serves is diverse; there is no monolithic demographic group. Gender is fairly evenly split, age is widely distributed, and race also has no one dominant group.

We also see this for the length of stay in the individual's prior living situation (Figure 7). The highest number of residents are at two extremes of the spectrum: one night or less, and one year or more.

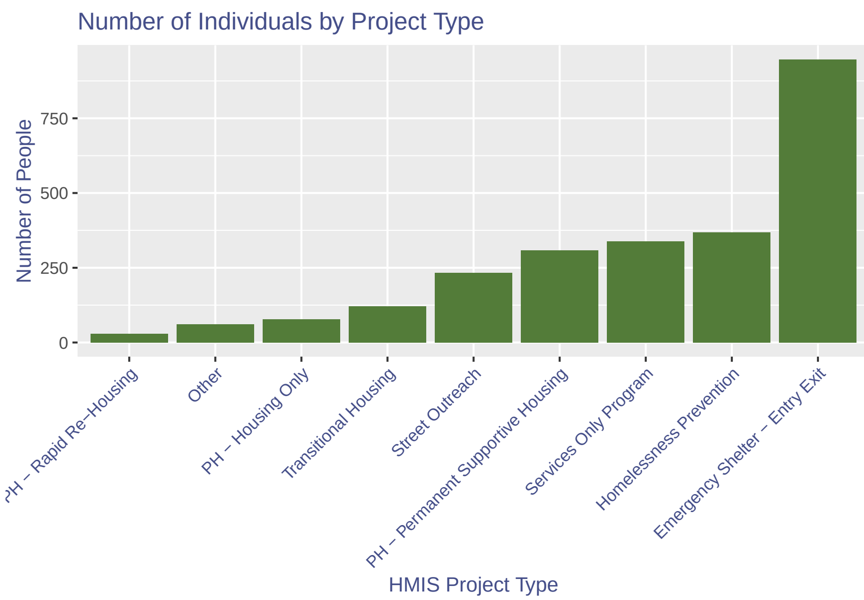

Figure 1: This graph displays the distribution of HMIS Residents by Project Type in increasing order.

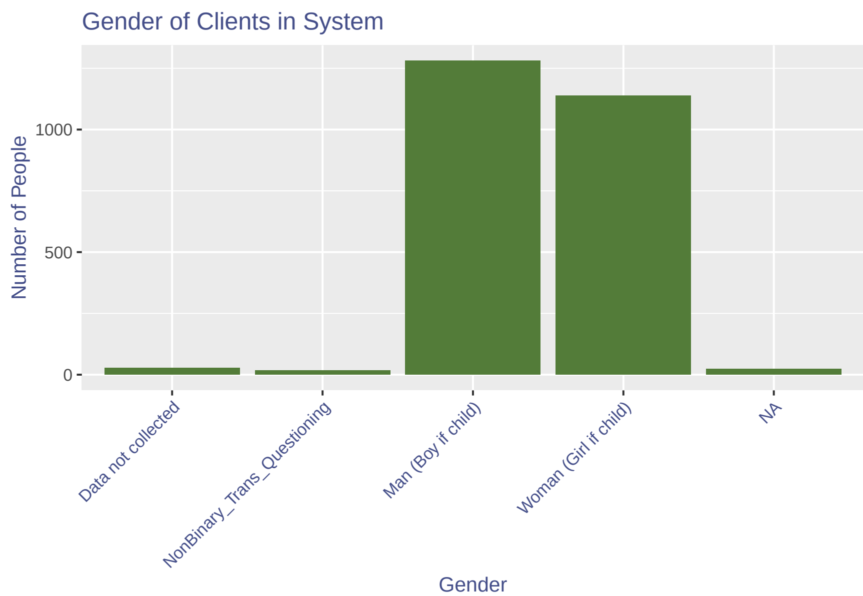

Figure 2: This graph shows the distribution of Resident Gender in HMIS. NA in this context represents the number of no inputs into the dataset for Gender.

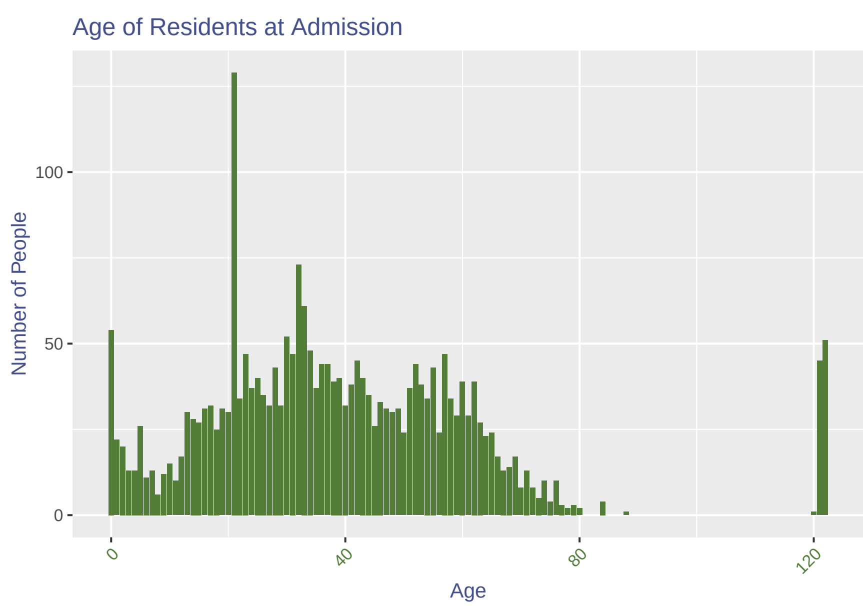

Figure 3: This barplot shows the Age of Residents at admission distributed across the entire dataset. The spike around 120 is inaccurate and likely due to case managers' method of inputting missing data, which is discussed at greater length in the Universal Data Elements and Missing Data Section.

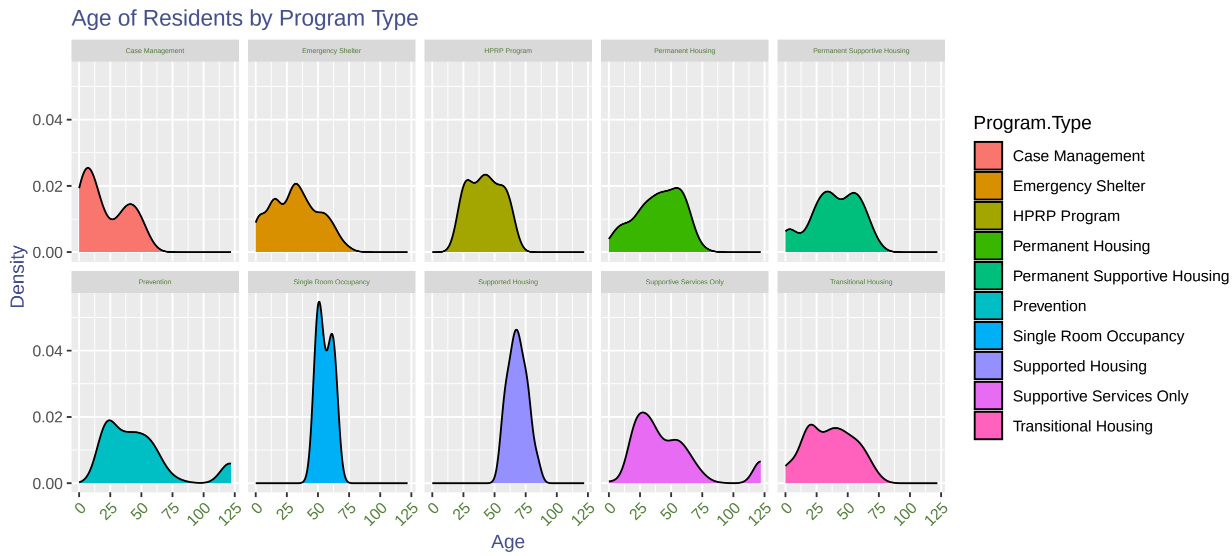

Figure 4: This density graph displays the distribution of Age at Admission for each Program Type in the dataset. The different curves provide a visual for what ages the data is distributed amongst in each Program Type.

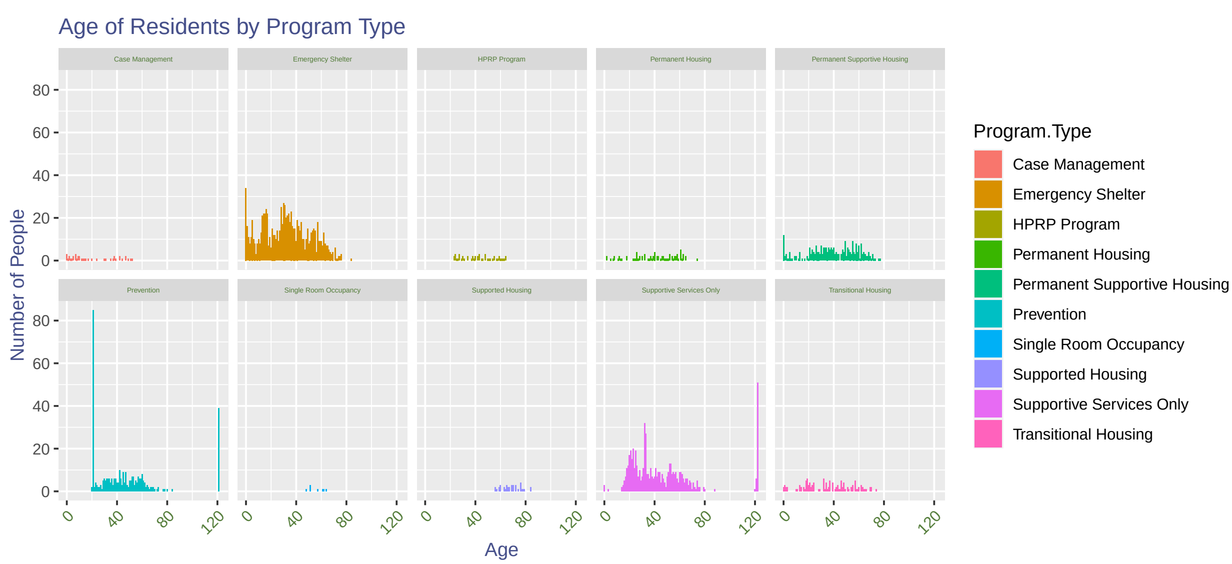

Figure 5: These histogram graphs display the distribution of Age at Admission for each Program Type in the dataset. You can notice trends by seeing spikes in certain ages for different program types.

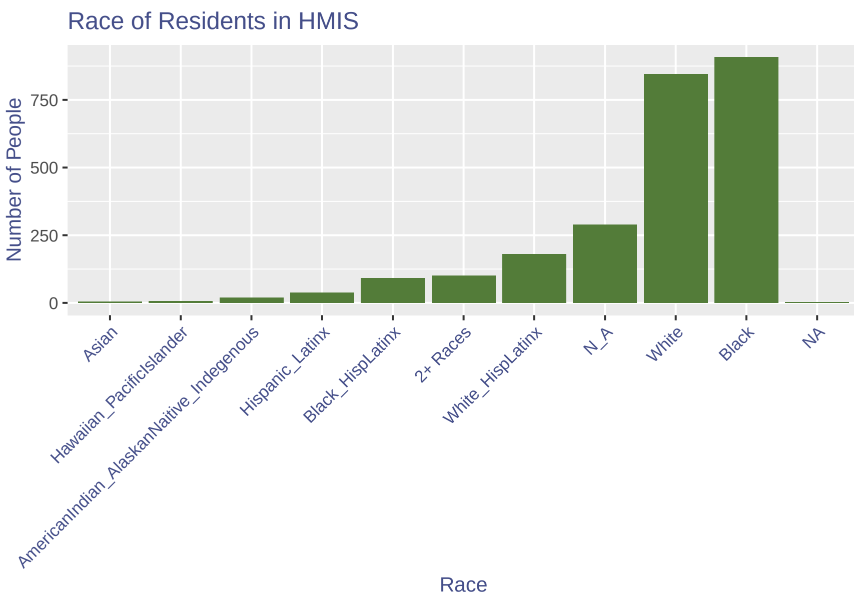

Figure 6: This graph displays the distribution of Race across the entire dataset. N-A in this context represents inputs such as clients refused to answer, data not collected, etc. Whereas NA represents the number of no inputs.

Figure 7: This graph shows the distribution of lengths of stay in the individual's living situation prior to admission. N/A in this context represents the number of times NA was inputted into the dataset for this variable, NA represents no inputs.

Trends in Length of Stay

Hudson River Housing offers many services to the community, including different types of affordable housing options, including emergency shelter, transitional housing, permanent supportive housing, and permanent affordable housing. In this section, we will look at one measure of the performance of these housing types: the length of program stay. This variable contains information about the number of days the individual has been in the system; this seems to be equivalent to either the number of days between admission and discharge, or if the individual has not yet been discharged, the difference between the admission date and the report generation date. As such, the reliability of this information depends on the accuracy of discharge data. For emergency shelters, individuals or households are discharged if they have not appeared at the shelter in the previous 48 hours.

We will also explore the question of whether there are certain characteristics related to a client's previous experience that might be related to, or predict, whether they stay in their program for longer or shorter. In particular, we were curious if there was any relationship between the length of program stay and the two following variables: prior transience, as measured by how long they lived in their prior living situation, and chronic homelessness.

Summarizing Length of Stay for Each Program Type

Table 1: Length of Stay (Days) by HRH Program

Program	Number of Clients	Mean (Days)	Median (Days)
HRH - Webster House	477	36.96	12.0
HRH - NYS ESG CV HP	350	908.55	911.5
HRH - Gannett House	310	143.53	123.0
HRH - Webster House Case Management	196	475.21	437.0
HRH - Pete's Place	127	205.81	184.0
HRH - SOP	107	139.09	32.0
HRH -River Haven ES	93	15.80	10.0
HRH - NYSSHP Affordable	78	2473.15	2304.0
HRH - Vet Zero	61	889.03	884.0
HRH - Hillcrest House	54	615.52	428.5
HRH - LaGrange House	52	1474.69	1211.0
HRH - RSP	48	288.73	287.0
HRH - AHEP	44	4.32	1.0
HRH - Liberty Station	41	1389.83	1238.0
HRH - Follow Through	40	701.20	582.0
HRH - EATS	38	763.82	808.0
HRH - Cannon Street	37	2965.54	2901.0
HRH - Fallkill Commons	37	946.81	1020.0
HRH - Elder Care Management	36	1075.56	1126.5
HRH - STEHP RRH	29	902.34	1004.0
HRH - United Way Follow-Through	29	920.62	904.0
HRH - Home Base I	27	2117.78	2786.0
HRH - May's House	23	98.74	64.0
HRH - Maximize	21	2951.05	1878.0
HRH - STEHP HP	19	854.21	930.0
HRH - River Haven TLC	15	249.73	208.0
HRH - 400 Maple Street	14	978.93	760.5
HRH - S Plus C	14	1533.36	971.0
HRH - 22 Lexington Avenue	13	856.46	1011.0

Program	Number of Clients	Mean (Days)	Median (Days)
HRH - Coach	10	1178.50	441.0
HRH - 472 Maple Street	8	1048.62	957.5
HRH - Noxon Street	8	3289.38	3426.5
HRH - 12 South Hamilton Street	7	1294.86	1225.0
HRH - Special Use Beds	7	2482.29	2032.0
HRH - 60 Catharine Street	5	1062.80	872.0
HRH - Forensic Supportive Housing	5	588.40	497.0
HRH - 81 Garden St.	3	1866.33	1714.0
HRH - RCE Special Use Beds	3	1081.33	1361.0
HRH - 107 N. Clinton	1	1261.00	1261.0

Table 2: Length of Stay (Days) by Program Type

Program.Type	Number of Clients	Mean (Days)	Median (Days)
Emergency Shelter	947	69.75	23.0
Supportive Services Only	594	476.15	396.0
Prevention	350	908.55	911.5
Permanent Supportive Housing	264	1273.89	932.5
Transitional Housing	121	939.40	646.0
Permanent Housing	78	2473.15	2304.0
HPRP Program	48	883.29	937.5
Case Management	40	701.20	582.0
Supported Housing	37	2965.54	2901.0
Single Room Occupancy	8	3289.38	3426.5

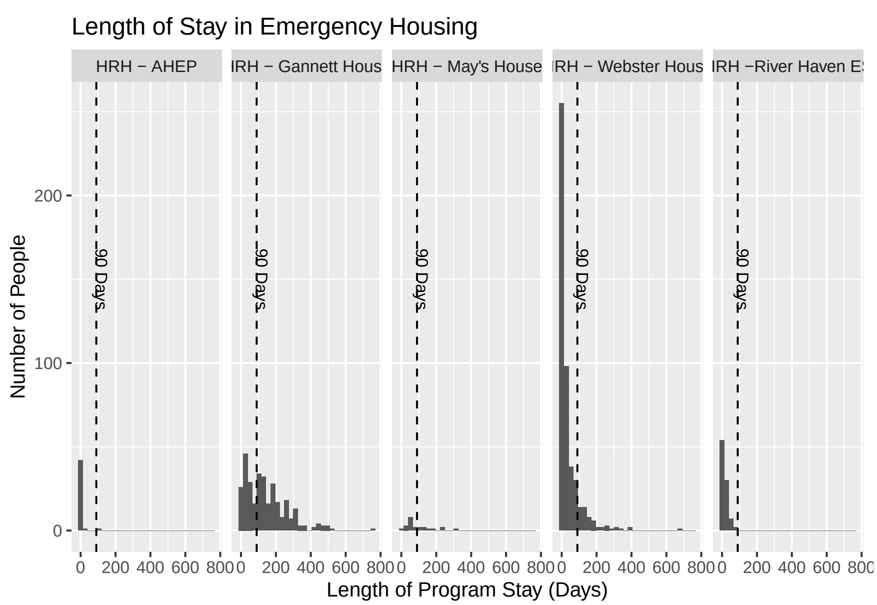

Figure 8: The length of program stays within specific Emergency Housing programs is shown, with a line at 90 days, representing the ideal time by which guest would be discharged

When we look at the length of stay for specific programs, the family-oriented services have higher median lengths of stay within emergency and transitional housing, compared to the services for individuals. Figure

Figure 9: The length of program stays within specific Transitional Housing programs is shown, with a line at 2 years, representing the HUD-specified maximum term of stay

8 shows the distribution of length of stay within emergency housing programs. The “ideal” time by which a guest would be moved out of emergency housing is 90 days, indicated by the dotted line in Figure 8. Transitional housing is intended to last for a maximum length of 24 months, as indicated by the dotted line in Figure 9. This difference between the distributions of length of stay in family programs versus in individual programs is in line with observations of the additional difficulty of finding affordable housing for families compared to individuals.

Tables 1 and 2 provide a breakdown of length of stay for each program and program type, sorted by the number of individuals receiving services from the program. One aspect to note is the discrepancy between the mean and median length of stay for many programs. When the mean is higher than the median, it indicates that the data is right-skewed; the values above than the median (the 50th percentile of the values) are more extreme and raise the mean. Within emergency shelter in Table 2, the mean is likely skewed by Gannett House, which provides emergency shelter for families. As seen in Table 1, the mean length of stay for Gannett House, roughly 143 days, is much larger than the mean for other emergency shelter programs such as Webster house, which has a mean length of stay of 37 days. This further highlights the shortfalls in availability of units in transitional and permanent housing for larger households.

Modeling Length of Stay for Each Program Type

We fit three linear regression models with a log-transformed outcome of length of program stay and predictors length of stay in prior living situation (a categorical variable) and chronic homelessness (a binary variable). We used three separate models given the differences in the scale and also meaning behind the “length of program stay” variable across programs (particularly given the specifics of the emergency housing discharge procedure) and the results can be seen in Table 3. For more model details and a more technical summary of the results, see the Appendix.

Interpretation of Table 3 is as follows. The first row shows the expected length of stay for the “reference” categories, i.e. someone who is not chronically homeless and has a length of stay prior living situation of over a year. The subsequent rows give the factor by which the expected length of stay increases relative to the value in row 1 as the relevant characteristic changes. For example, based on this model, someone in emergency housing who is chronically homeless and prior living situation greater than a year has expected length of stay

$12.05 \times 0.77 \approx 9$ days. Someone in permanent supportive housing who is chronically homeless and lived in their prior living situation for less than a week has an expected length of stay of about $1016.52 \times 1.33 \times 0.46 \approx 622$ days.

The most notable results relate to transitional and permanent supportive housing. In transitional housing, the model estimates that residents who lived in their prior living situation for over a year have longer expected lengths of stay compared to those whose length of prior living situation, after adjusting for chronic homelessness status. In permanent supportive housing, the model estimate shows that the expected length of stay for an individual who is chronically homeless at admission is about half the length of stay of that of a non-chronically homeless individual. These two results are statistically significant for $\alpha = 0.10$; i.e. we are reasonably confident that there is this true difference.

Table 3: Model for Length of Stay in Housing (Days)

	Model Estimate by Housing Type		
	Emergency	Transitional	Permanent Supportive
Expected Stay (Days): Non-chronically-homeless, prior living situation > 1 year	12.05	741.11	1016.52
Times Longer for Prior Living Situation 1 month-1 year	1.72	0.64	0.91
Times Longer for Prior Living Situation 1 week-1 month	1.08	0.31	1.13
Times Longer for Prior Living Situation less than 1 week	1.25	0.42	1.33
Times Longer for Chronically Homeless	0.77	1.21	0.46

Data Completeness and Quality

Universal Data Elements and Missing Data

One of our goals in this analysis was to explore patterns among incomplete data from 2023. In this analysis, we focused primarily on the Universal Data Elements (UDEs). Because collection of Universal Data Elements is required, we look at instances of missing values for these variables.

Our main focus is examining the incomplete data centered around the universal data elements. Because we did not have access to identifying information such as birthdays, SSNs and other UDEs, we examined a subset of these variables. These variables, as shown in Table 5, have varying degrees of missingness. Additionally, the separation of this missingness into "data not inputted" and "data not collected" gives more insight as to where these missing entries are coming from. "Data not inputted" denotes a question that was left blank on an intake form, while "data not collected" encompasses the responses "data not collected", "Client refused", "Client doesn't know", or "N/A". The significance of the missingness in Table 5 also varies between the different variables. As we discussed in our meetings, the continuum of care variable has a high degree of missingness, but this is likely deliberate and is being used to help residents find housing more efficiently. A visualization of the distribution of missing data by program and program type is displayed in Figure 11. On the other hand, more diligent collection of the gender, disabling condition, and veteran status variables would lead to more complete data. This would allow for the data to represent a more complete picture of the population that HRH serves.

One of the universal data elements not included in Table 5 was age. Because a resident's birthday cannot be skipped at admission, we have responses from all residents about age in the dataset. In responses to the age variable, we saw that 97 people were listed as being between 120-122 years old. In conversations with HRH, we determined that this was likely a method used by case managers in situations where the date of birth was not obtained from the resident. In the data, we saw that these people aged 120+ came from only 3 programs: "NYS ESG CV HP", Pete's Place, and "SOP", which is shown in Figure 10. Because we only saw 97 instances of ages that we could definitively identify as not being plausible, and these entries came from only 3 programs, it is impossible to say if other locations are using different standards for missing DOB. With this in mind, we recommend that HRH implements a standard protocol for inputting a missing DOB into the HMIS system. This date should be standardized across all programs in order to maintain consistency in the data, and should be easily identifiable as a missing entry (eg. age 120+). Although our analysis mainly focused on age, a protocol for standardization of missing variables should be implemented across any required variables where this problem could arise.

Other Data Recommendations

Although we focused mainly on the universal data elements in our analysis, we also wanted to explore which variables that were not being collected might have a correlation with length of stay. The goal of this analysis was to identify a handful of variables that could be used as predictors in future modeling of length of stay. In these variables, we saw a difference in the median lengths of stay between different responses to the variables, however these variables also had significant missingness. Table 6 displays the 4 variables we identified from this analysis, as well as the percentage that these variables were missing in the dataset. We recommend that these variables be prioritized in future data collection.

Another recommendation is to define and record an additional variable for positive versus negative outcomes at discharge, or some other time point. This would be somewhat subjective, but defining outcomes would allow for calculation of the rates of these positive/negative outcomes, and comparisons can be made to inform future policies and procedures.

Additionally, it is worth noting that 15 variables had no responses, meaning that nobody had answered them in the admission of any resident. These variables are listed in Table 3.

Table 4: Table of variables that were completely missing (no data inputted for any resident). These were variables that we excluded from the dataset for our analysis.

Variable Name
variable.names(NA__vars)
Current.Living.Situation..Admission.
Does.Individual.or.family.have.resources.or.support.networks.to.obtain.other.permanent.housing...Admission.
Has.a.subsequent.residence.been.identified...Admission.
Has.the.client.had.a.lease.or.ownership.interest.in.a.permanent.housing.unit.in.the.last.60.days...Admission.
Has.the.client.moved.2.or.more.times.in.the.last.60.days...Admission.
Is.client.going.to.leave.their.current.living.situation.within.14.days...Admission.
Location.Details..Admission.
Name.of.Affiliated.Residential.Project
Number.of.Months.Ward.of.Juvenile.Justice.System
Preferred.Language.Different
Referral.From.Hospital
Rental.Subsidy.Type..Admission.
RHY.Number.of.times.approached.by.outreach.prior.to.entering.the.project
School
Veteran.Unit.Number

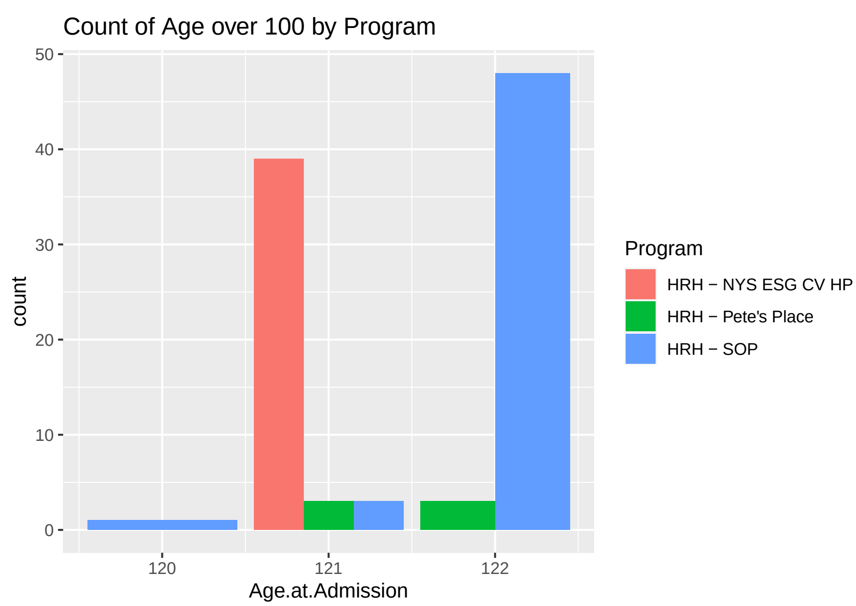

Figure 10: Age Greater than 100

Table 5: Number of Missing Values Within Selected Universal Data Element Variables

variable	No__Data__Inputted	Data__Not__Collected	Total__Missing
Gender..HMIS.	24	27	51
Disabling.Condition	0	447	447

variable	No_Data_Inputted	Data_Not_Collected	Total_Missing
Race..HMIS.	3	0	3
Veteran.Status	0	370	370
Enrollment.CoC..Admission.	508	0	508

Table 6: Recommended Variables with Percent of Observations Missing

Variable	Percent.Missing
On.the.night.before.did.you.stay.on.the.streets..ES.or.SH. . . Admission.	95.90
Total.number.of.months.homeless.on.the.street..in.ES..or.SH.in.the.past.three.years..Admission.	50.74
Sexual.Orientation	91.64
Mental.Health.Disorder..Admission	23.04

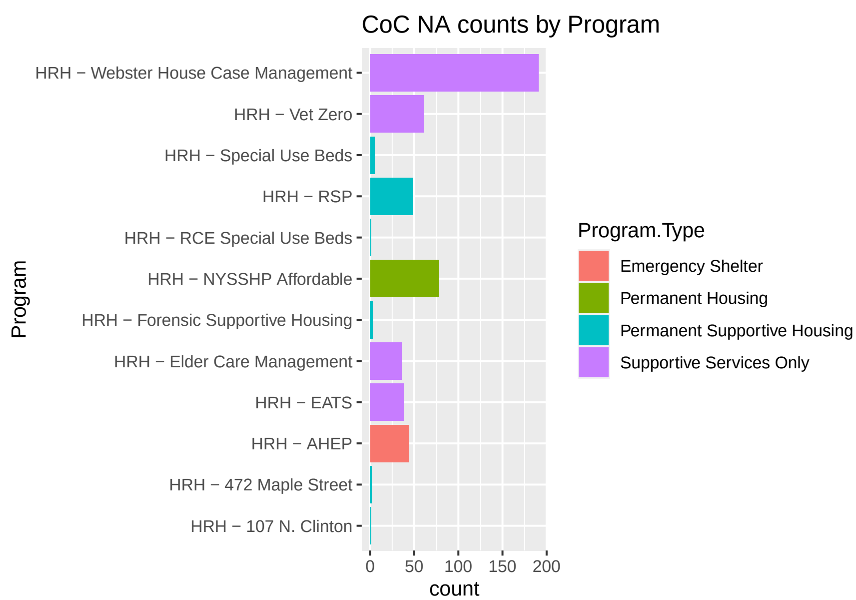

Figure 11: CoC Enrollment Missing Entries by Program

Conclusion

Data analysis can provide quantitative support for observations and experience of those who work in the system on a daily basis. Data and data visualization can also serve as tools to communicate with the general public. Lastly, standardization of data collection going forward can help guide the transition towards being more data driven

Acknowledgements

Thank you to Hudson River Housing for the opportunity to work with them and for trusting us with this project. Special thank you for Javier Gomez for lending his expertise and time to this collaborative project.

Appendix

Model Details

The data for the model was restricted to include only individuals who were marked as the primary client (one representative individual per household), due to dependence, or similar values for many variables, within families. Missing data was treated as missing at random and simply excluded, but this is likely not true. As such, these results may be biased.
A printout of the linear regression model results is below.

Emergency Housing:

```
##
## Call:
## lm(formula = log(Length.of.Program.Stay) ~ Length.Prior.Living.Situation +
##     Chronically.Homeless..Admission., data = filter(model_selfonly,
##     HMIS.Project.Type == "Emergency Shelter - Entry Exit"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0300 -1.6125  0.1221  1.4943  4.0516
##
## Coefficients:
##                                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)                              2.48931    0.17560  14.176   <2e-16
## Length.Prior.Living.Situationmoto1yr     0.54072    0.25109   2.154   0.0316
## Length.Prior.Living.Situationweekto1mo   0.07617    0.33150   0.230   0.8183
## Length.Prior.Living.Situationlessthan1week  0.22175    0.19302   1.149   0.2510
## Chronically.Homeless..Admission.Yes     -0.26138    0.35932  -0.727   0.4672
##
## (Intercept)                             ***
## Length.Prior.Living.Situationmoto1yr    *
## Length.Prior.Living.Situationweekto1mo
## Length.Prior.Living.Situationlessthan1week
## Chronically.Homeless..Admission.Yes
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.795 on 718 degrees of freedom
```

```
##   (34 observations deleted due to missingness)
## Multiple R-squared:  0.008204,   Adjusted R-squared:  0.002679
## F-statistic: 1.485 on 4 and 718 DF,  p-value: 0.205
```

Transitional Housing:

```
##
## Call:
## lm(formula = log(Length.of.Program.Stay) ~ Length.Prior.Living.Situation +
##     Chronically.Homeless..Admission., data = filter(model_selfonly,
##     HMIS.Project.Type == "Transitional Housing"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2670 -0.7729 -0.0326  0.8030  2.5194
##
## Coefficients:
##                                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                  6.6081     0.2230  29.635   <2e-16
## Length.Prior.Living.Situationmoto1yr        -0.4508     0.2619  -1.721   0.0884
## Length.Prior.Living.Situationweekto1mo      -1.1807     0.5530  -2.135   0.0353
## Length.Prior.Living.Situationlessthan1week  -0.8678     0.4076  -2.129   0.0358
## Chronically.Homeless..Admission.Yes          0.1888     0.3535   0.534   0.5946
##
## (Intercept)                                ***
## Length.Prior.Living.Situationmoto1yr       .
## Length.Prior.Living.Situationweekto1mo     *
## Length.Prior.Living.Situationlessthan1week *
## Chronically.Homeless..Admission.Yes
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.132 on 96 degrees of freedom
## Multiple R-squared:  0.08286,    Adjusted R-squared:  0.04464
## F-statistic: 2.168 on 4 and 96 DF,  p-value: 0.07833
```

Permanent Supportive Housing:

```
##
## Call:
## lm(formula = log(Length.of.Program.Stay) ~ Length.Prior.Living.Situation +
##     Chronically.Homeless..Admission., data = filter(model_selfonly,
##     HMIS.Project.Type == "PH - Permanent Supportive Housing"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7461 -0.8056  0.1177  1.0483  1.8869
##
## Coefficients:
##                                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                  6.9241     0.1037  66.791  < 2e-16
## Length.Prior.Living.Situationmoto1yr        -0.0939     0.1422  -0.660 0.509581
## Length.Prior.Living.Situationweekto1mo       0.1225     0.3216   0.381 0.703712
```

```
## Length.Prior.Living.Situationlessthan1week   0.2884     0.3217   0.897 0.370797
## Chronically.Homeless..Admission.Yes         -0.7690     0.2066  -3.723 0.000244
##
## (Intercept)                                ***
## Length.Prior.Living.Situationmoto1yr
## Length.Prior.Living.Situationweekto1mo
## Length.Prior.Living.Situationlessthan1week
## Chronically.Homeless..Admission.Yes        ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.06 on 247 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.05801,    Adjusted R-squared:  0.04276
## F-statistic: 3.803 on 4 and 247 DF,  p-value: 0.005087
```