DSP 539 Final: Male Sex Workers in Rhode Island

Emiton Alves

5/10/2019

Background Information

Internationally, male sex workers have been shown to HIV prevalence rates of up to 50%. Although HIV prevalence estimates among this population are not this high in the United States, they are still significantly higher when compared to other high-risk groups such as men having sex with men (7.3% vs 1.1%). These high rates can be attributed to the fact that many of these men report engaging in unprotected sex with male clientele as well as sharing needles with others when injecting drugs – behaviors that have both been shown to significantly increase risk of HIV acquisition. Despite being at higher risk of HIV acquisition, male sex workers will often forego preventative care and treatment services because of substantial social stigma. It is important to note that many of these men and their predominantly male clientele identify as heterosexual or bisexual – many have wives, girlfriends or other female partners with whom they frequently engage in unprotected sex2,5. As such, this group, directly or indirectly, can act as a vector of HIV transmission to several different populations2-5. For these reasons, identifying factors that increase risky behaviors among male sex workers is an important step in reducing the incidence of HIV. With this information, interventions can be developed to address the underlying causes of risky behaviors. In 2008, a RI-based agency called Project Weber (now Weber-Renew) was created by two former male sex workers who saw a need to increase access to harm-reduction services to this high-risk population. The organization initially started with a small team of volunteers conducting street-based outreach, but in 2013, it received state funding to open up a drop-in center. Run by peer educators, the drop-in center provides male sex workers a safe space where they could receive harm-reduction services such as risk-reduction counseling, sexually transmitted infection (STI) testing and needle exchange.

Data Description

The data used for this analysis were collected as part of a survey completed by clients of Project Weber. All surveys were conducted by trained peer-educators at Project Weber's drop-in center from February to December of 2014. Before analysis was conducted, the data were stripped of any identifiable information by the organization. See Appendix 1 for Client Survey. The variables included in this analysis were the following:

Age

- Type: Numerical Variable (Interval)
- Units: Years

Education Level

- Type: Categorical Variable (Ordinal)
- Levels: Less than high school, High school or GED, Some college, College degree, Some graduate work (no degree to date), Graduate/professional degree

Recent Unstable Housing (Past 6 Months)

- Type: Categorical Variable (Nominal)
- Levels: Yes, No

Ever Incarcerated

- Type: Categorical Variable (Nominal)
- Levels: Yes, No

Recent Access to Healthcare Services (Past 6 Months)

- Type: Categorical Variable (Nominal)
- Levels: Yes, No

Recent Violence (Emotional, Physical or Sexual) Experienced (Past Year)

- Type: Categorical Variable (Nominal)
- Levels: Yes, No

Number of Years Doing Sex Work

- Type: Numerical Variable (Ratio)
- Units: Years

Recent Drug Use (Past 6 Months)

- Type: Categorical Variable (Nominal)
- Levels: Yes, No

Number of Female Partners (Past Month)

- Type: Numerical Variable (Ratio)
- Units: Persons

Number of Male Partners (Past Month)

- Type: Numerical Variable (Ratio)
- Units: Persons

Recent Unprotected Sex with Male (Past Month)

- Type: Categorical Variable (Nominal)
- Levels: Yes, No

**Recent Unprotected Sex with Female (Past Month)**

- Type: Categorical Variable (Nominal)
- Levels: Yes, No

Data Pre-processing

The survey consists of 44 questions and the information was collected by trained professionals. Some the of 44 questions ask for similar information such as health care usage or current/past living situation. This produces features which are highly collinear with other variables within the dataset. Additionally, there were only 140 surveys completed. Given the sample size and the occurrence of highly collinear features, it was decided that a limit of about 12 features should be used for the learning process. This number was chosen by taking the square root of the sample size. It is a general rule of thumb when working with highly correlated data. Out of the 140 samples, 23 had to be dropped due to missing values. This leaves a dataset of 117 samples. Following the previous rule of thumb, there should be an upper limit of about 11 features on the reduced dataset. This feature limit is not a strict one. The number of features can be increased if

deemed necessary. The features used in the analysis have been described in the data description section. Of these features some were chosen out of a group of other similar questions. For instance, question 8 asked the client whether or not they has access to stable housing within the past six months. There were 3 other questions which also asked about housing, but question 8 seemed to be the most prevalent housing question to use for this analysis. Similar decisions were made about every other question chosen as a feature for the analysis. Given that the data was collected by trained professionals, the main factor affecting the dataset was missing values. After removing the samples which had missing values, it is believed that the remaining data is prepared for a thorough analysis.

Data Analysis
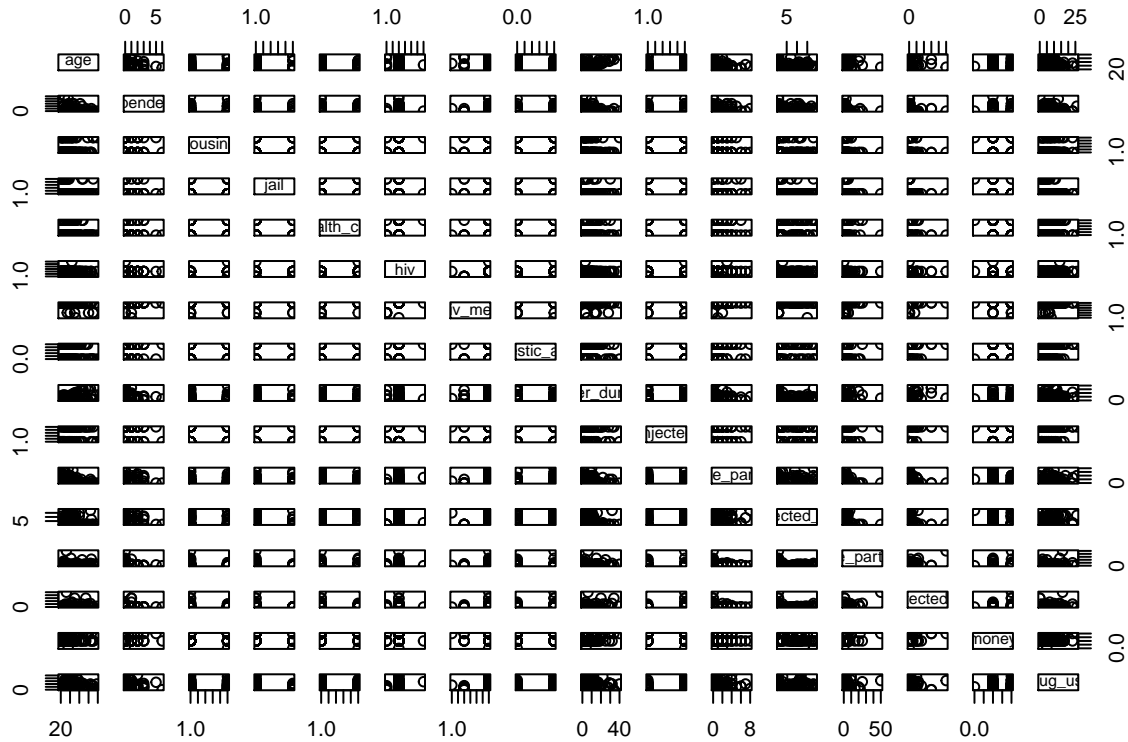
The figure above is a pairwise correlation matrix. It displays all variables plotted against one another for instance the second box on the first line is age versus dependents. A pairwise plot is a great way to visually inspect if there is any correlation between two given variables. In the case of this dataset there is a clear correlation between age and career duration. This seems to be an intuitive find. Upon first inspection there does not too be too many variables which seem to have a linear correlation.

The next focus of the anaylsis is simply to see and understand the spread and shape of the data.

## Figure 2: Housing in the Past 6 Months
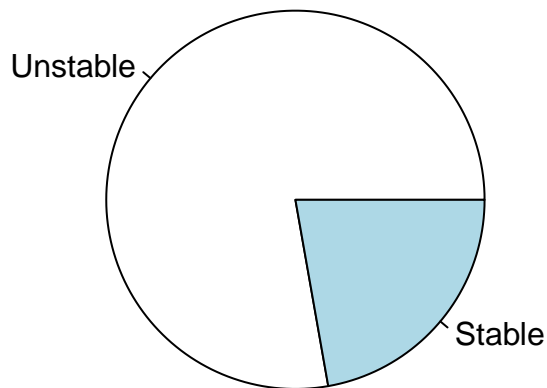
The pie chart above shows that most of the participants within the study did not have recent access to stable housing. Over 75% of people in the study reported having to rely on different friends, family, or sometimes even on customers to provide them with a place to sleep.

Figure 3: Client Age
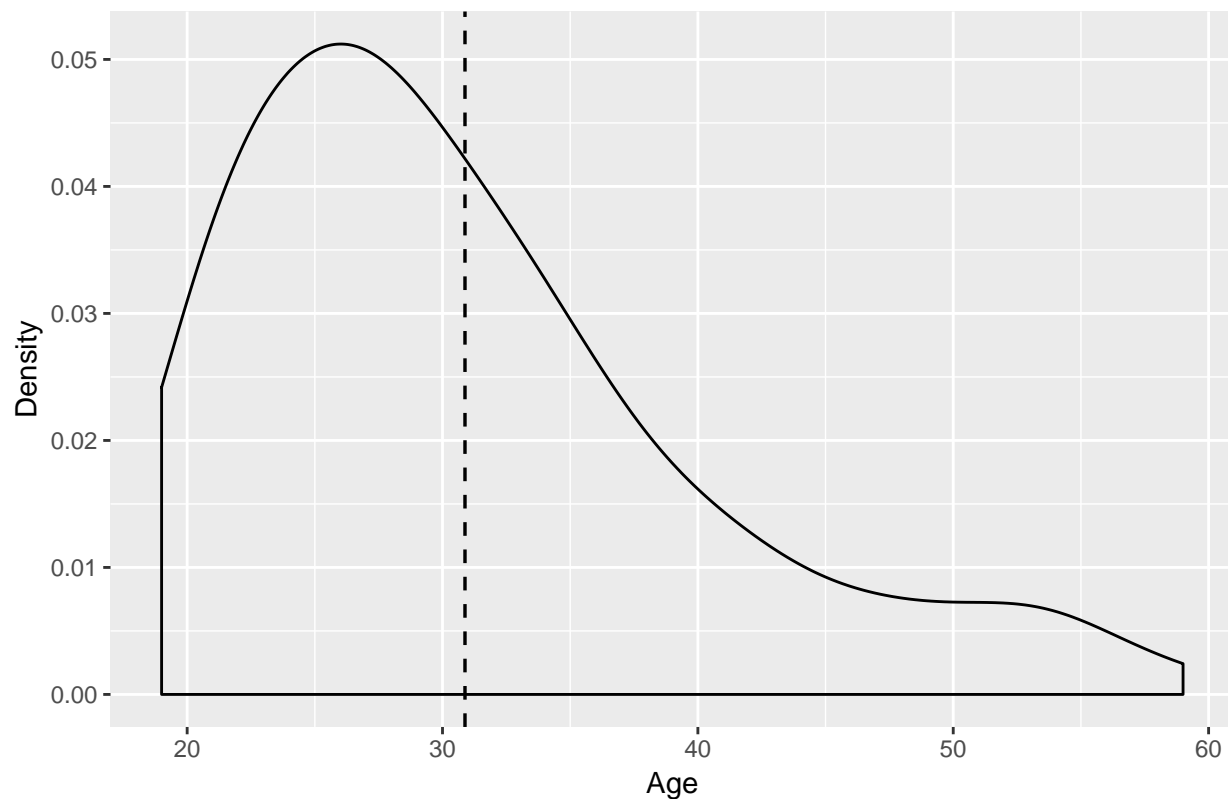
This chart displays the values for the ages of participants in the study. Here it is observed that a large portion of the clients are between ages 20-35 with the average age being just above 30. There was a large spread between with the minimum ages and the maximum age. The youngest client was 18 and the oldest client was 59, a difference of 41 years.

Figure 4: Time in Occupation
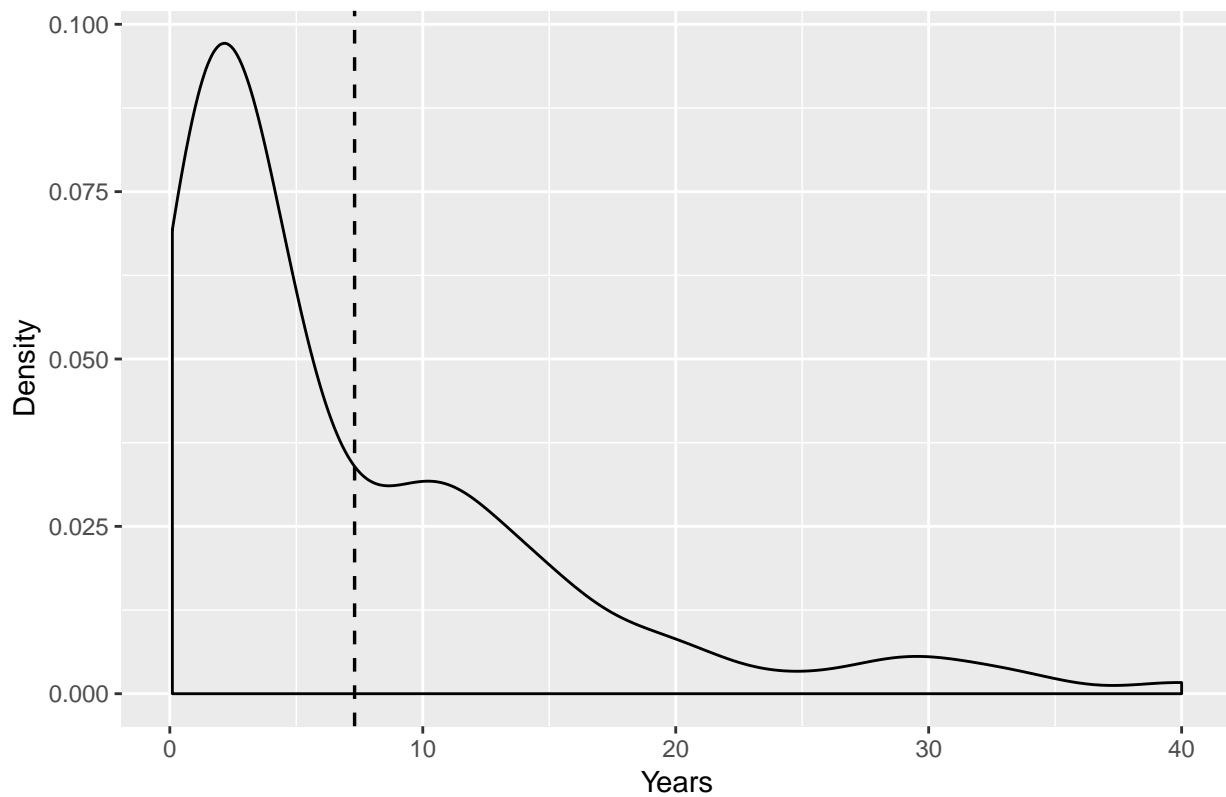
This chart displays the career duration of clients within the study. It can be seen that most clients had a duration between 0-6 years. The average length of a clients career had been 7 years. There is also some outliers observed within the data. Although the average career length is 7 years, there are of few cases of 30 and 40 year careers lengths. This skews the data towards an artificially high average.
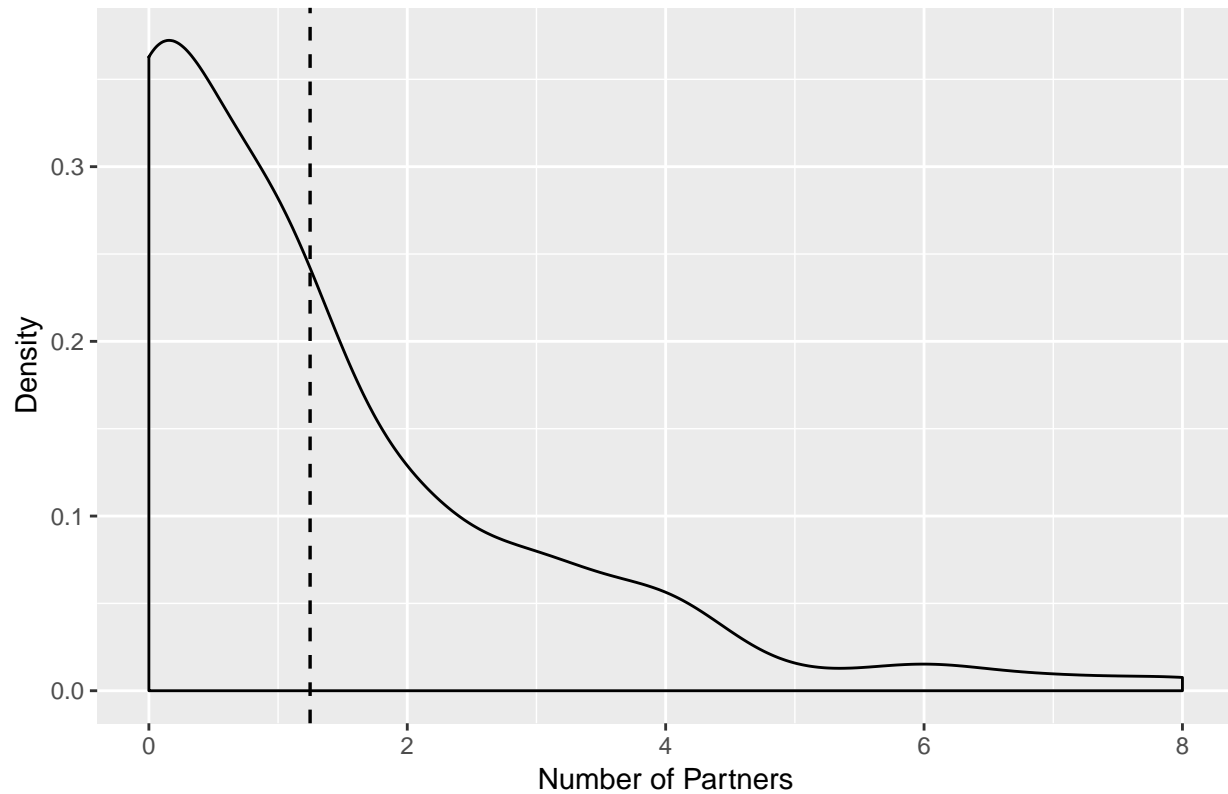
Figure 5: Female Partners

Figure 5 displays the reported number of female sex partners for each client. Here it is seen that the average female partner is only 1. Most clients did not report serving many female customers. Although the maximum number of female partners was 8, a large portion of clients had no more than 2 female partners.
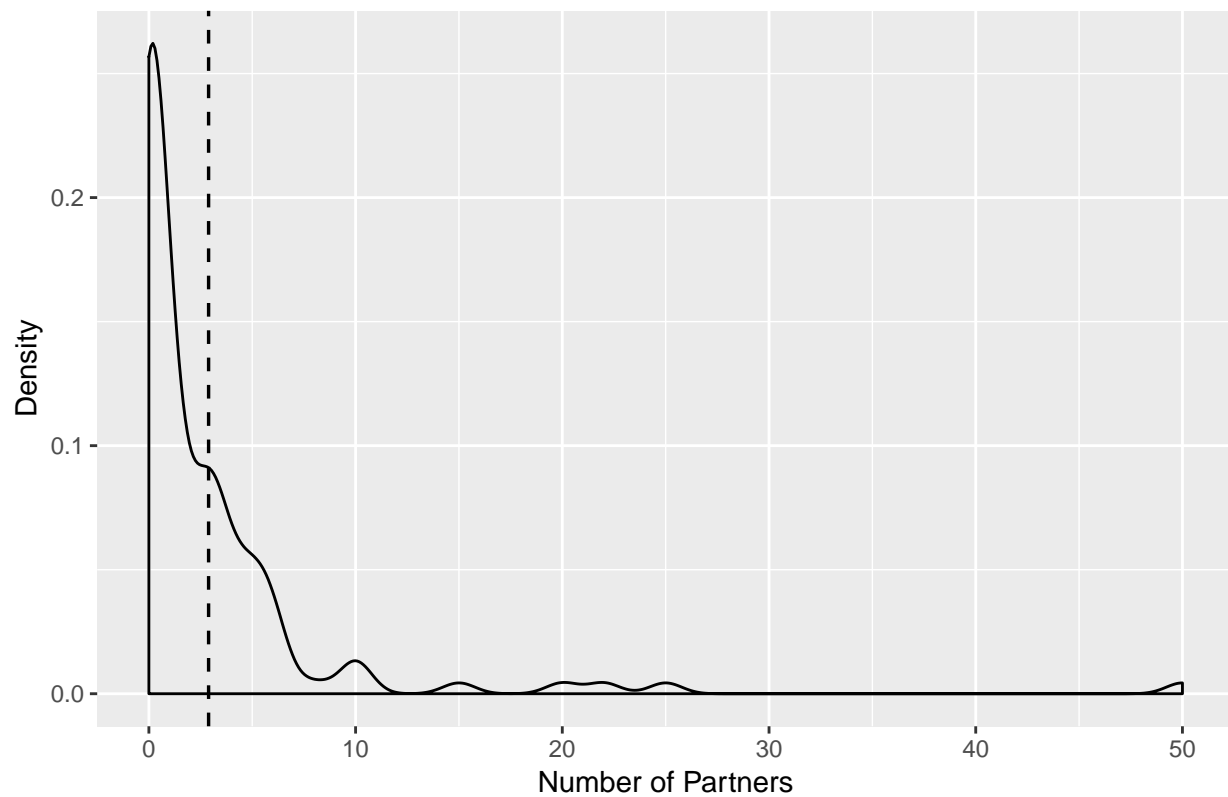
Figure 6: Male Partners

Figure 6 displays the number of male sex partners that each client reported. Here it is seen that there is a much greater spread of the number of partners had. A large portion of clients reported having less than 5 partners. The average number of male sex partners was 3. Some outliers had reported as much as having had 50 partners.

```
## Scale for 'fill' is already present. Adding another scale for 'fill',
## which will replace the existing scale.
```

Figure 7: Numerical Correlation
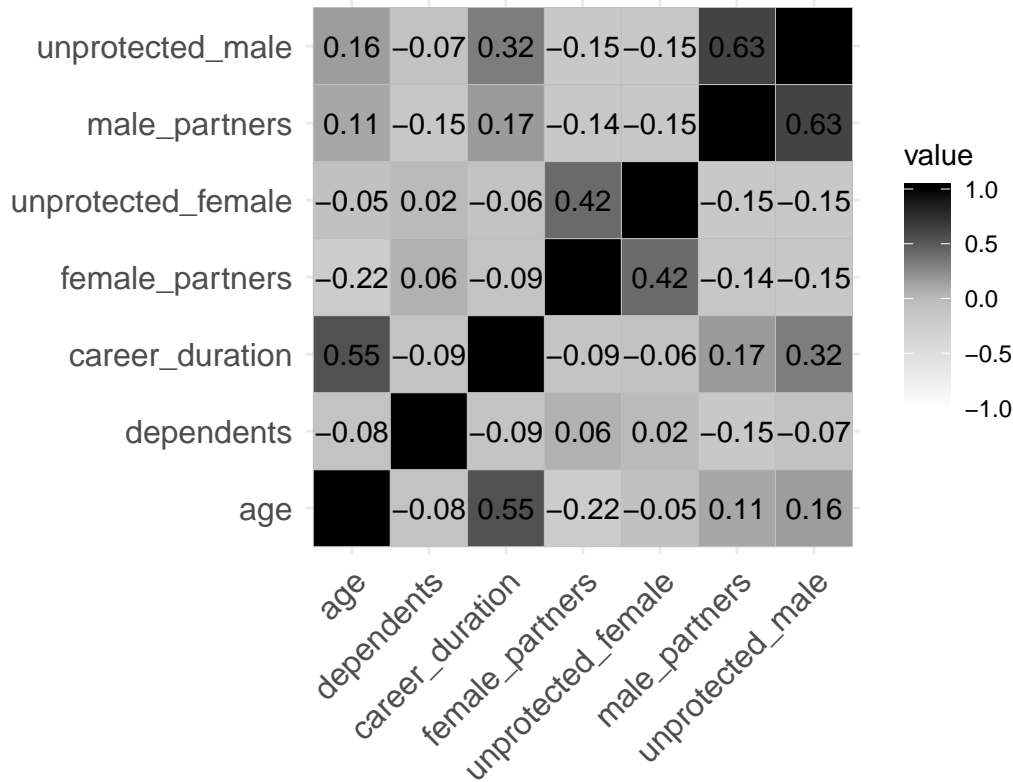
The correlation matrix in figure 7 displays the numerical correlation between variables which are currently of highest interest. Here it is much easier to differentiate between the correlation and each variable's effect on other variables. Here is it also observed that age and career duration have a positive correlation which was discussed earlier.

A linear model was applied to the dataset in order to infer the effect of given variables on the probability of contracting HIV. Please run "summary(hiv_model)" within the "linear_model" code block in order to see the raw output of the model. In this model, it was observed that in this model there was no significant predictor at the 5% level for HIV contraction. It was also observed that clients who contracted HIV tended to be older than HIV negative clients.

Conclusion

This data provides a baseline assessment of risk, but it does not show the effect of interventions such as regular STI testing, needle exchange or risk-reduction counseling. Follow-up data on the utilization of these services and a reassessment of risk would provide the researchers with an opportunity to measure the effect of such interventions. Additional analysis must be carried out in order to determine which variables will help best with the model. As mentioned before, questions on the survey were chosen to reduce collinearity. The topics of housing, schooling, jail, sexual partners, were only represented once. Variables such as drug use for different drugs were collapsed into one variable. The current variables were chosen, because they contained data with more normal distributions. Further tests will be carried to isolate and variables at in order to determine correlation and effective use.