Drug and Substance Abuse Treatment

Violet Allen-Glass, Isis Ramirez, Kaleigh Riggs, Jinchen Ye

Introduction

Primary Data

The main data set we are used cames from the Center for Behavioral Health Statistics and Quality Substance Abuse and Mental Health Services Administration. It is the Treatment Episode Data Set- Admissions for 2016. It includes 1,699,261 observations and 62 variables, each of these representing a person that was admitted for substance abuse to a publicly funded facility. Omitted from this data were the states Georgia and Oregon, because they did not report sufficient data in 2016.

Overall our objective was to identify what kind of people were admitted to these public facilities, including their lifestyles, ages, and specific drug substance abuse. Additionally we looked at many factors across race to identify any racial disparities in access to care. To achieve this we wanted to find any trends in age admitted by race, and age admitted by gender. We looked for trends in services rendered or referral sources among the patient's employment status, and further we looked at referral sources by race. For both employment status and wait time until treatment we used chi-squared hypothesis tests to recognize any significant differences by race.

Below we list the coverage for the variables used from our main data set.

Variable	Variable Description	Coverage %
CASEID	Case ID	100.00
RACE	Race	97.58
AGE	Age	100.00
EMPLOY	Employment Status	94.88
PSOURCE	Referral Source	95.35
DETNLF	Detailed Not In Labor Force	30.42
SERVICES	Type of Treatment	100.00
SUB1	Primary Substance Use	99.05
SEX	Gender	99.95
DAYWAIT	Days Waited For Treatment	55.44
YEAR	Year	100.00

Secondary Data

The secondary data we used came from the National Health Interview Survey 2016 data released by the Center for Disease Control Prevention's National Center for Health Statistics. The specific dataset we used is for a sample of adults from this survey. It has 33,028 observations and 805 variables, which provide details of their physical and mental wellbeing.

With this data set we looked at the mental health of substance abusers by race.

Tertiary Data

Our tertiary data set was scraped from a 2017 Reddit post with 1,083 comments in which drug users discussed what drugs they tried and how those substances made them feel.

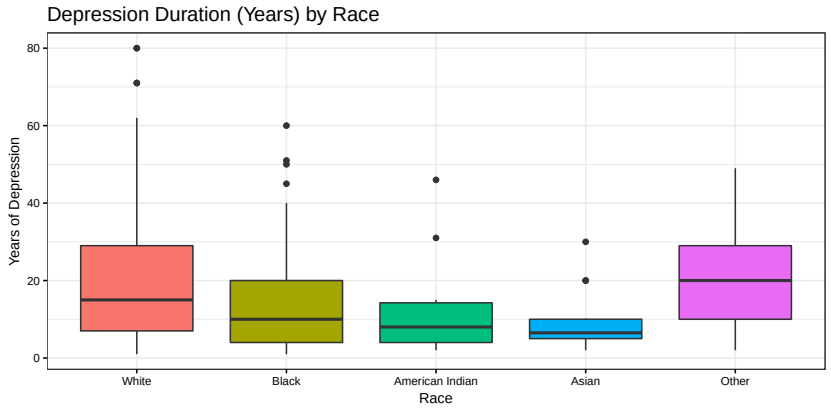
In order to identify what kind of experience users had with different drugs, we compared the percent of positive and negative comments for the most frequently mentioned drugs via text mining.

Mental Health & Substance Abuse

Using our secondary data source, we looked at how many interviewees who struggled with substance abuse also struggled with depression. We see that of those that indicated having a substance abuse issue, about 39% also struggled with depression.

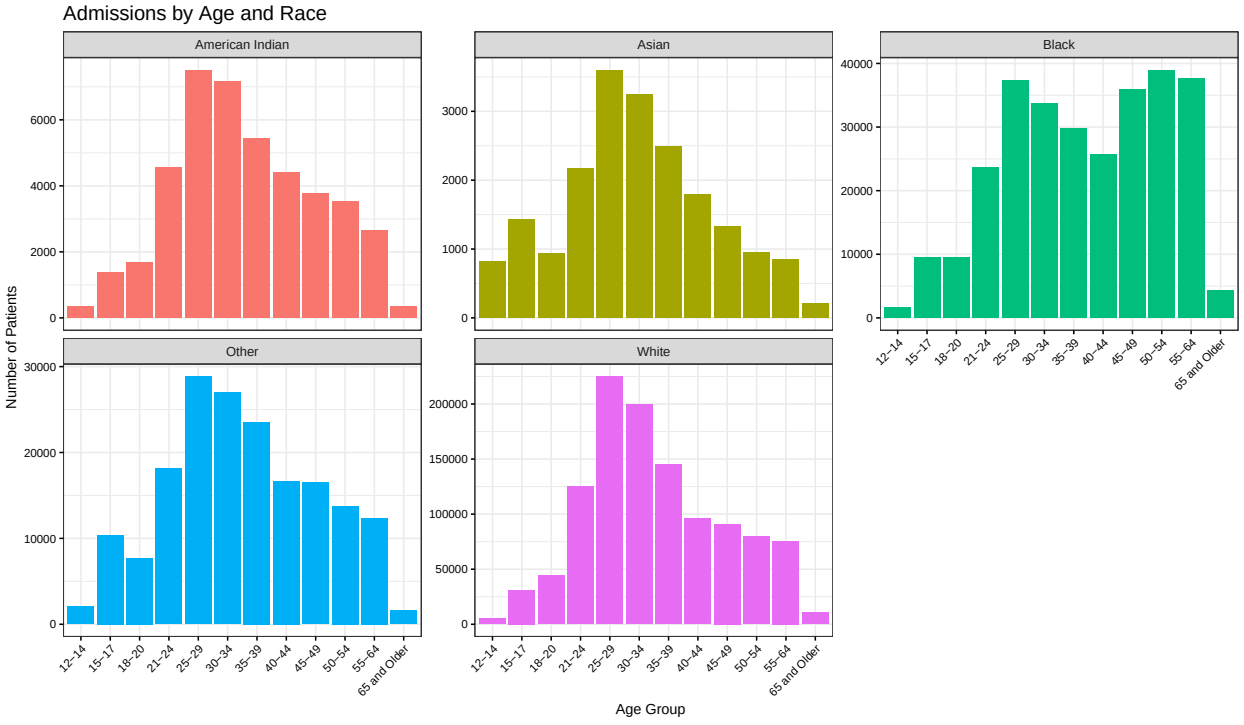
For that 39%, we looked at how long they suffered from depression. White and Other races seemed to suffer from depression the longest. The median for the Other race group was the highest with a duration of 20 years.

Depression Status	Proportion
Depression	0.39
No depression	0.61



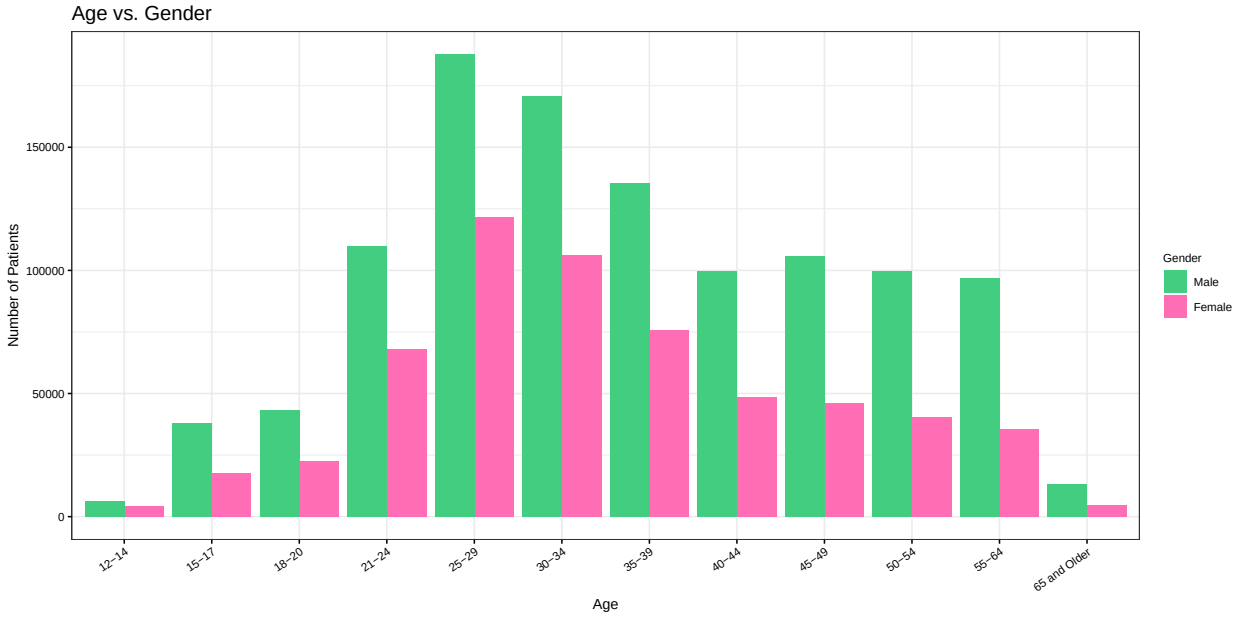
Age

We searched for distinct trends in age of admissions by race. While most races seem approximately bell curved, peaking in ages 25-29, the age that black patients are admitted does not taper down in older years, but instead seems bi-modal with a second peak at ages 50-54.



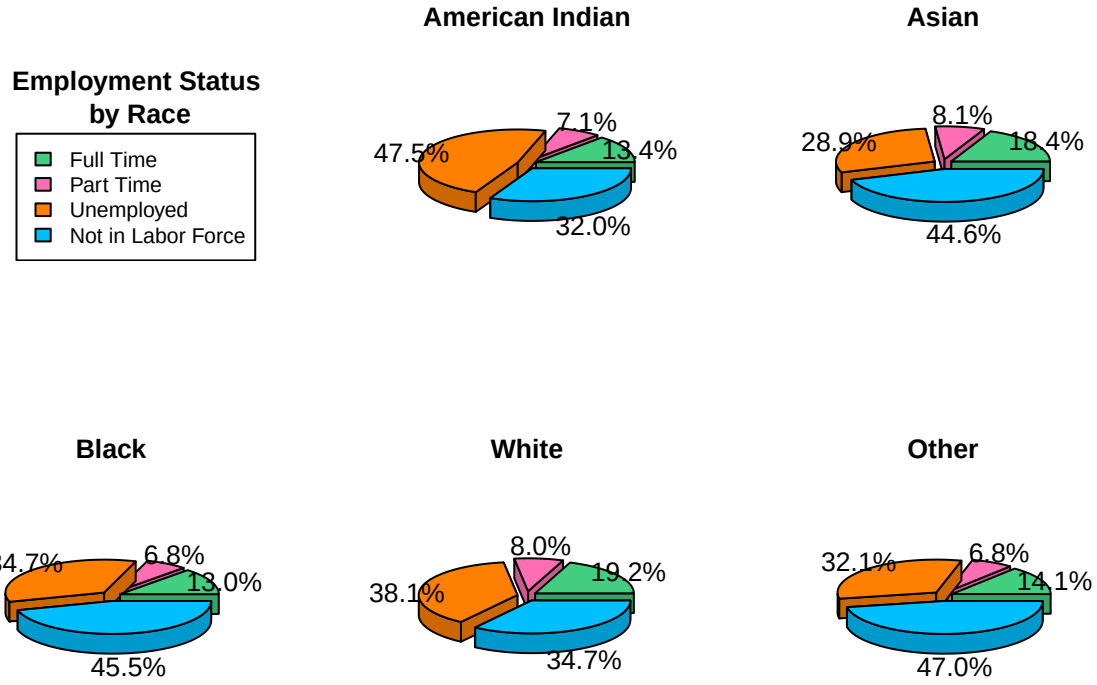
The bar graph below shows the number of admissions by age group and gender. We see that the majority of patients admitted are between the ages of 25-29, and that across all age groups, the majority of patients

admitted are male.



Employment Status

We looked at the employment status of patients by race. Our goal here was to see whether day-to-day life had any affect on whether someone was more likely to seek treatment for substance abuse. We found that unemployed and not in labor force are the majority of admittees across all races. It is interesting that across the board, part time employees were the smallest percent admitted, even less than full time employees. Not in Labor force includes homemakers, students, retirees, disabled, residents of institution.



Our next step was to test for a significant difference among employment status for the defined race groups using a chi-squared hypothesis test.

Chi-Square Hypothesis Test

The assumptions of the test, all of which were met, were that:

1. We were analyzing count data
2. Variables were mutually exclusive
3. Each subject contributed to only one cell
4. Independence
5. Categorical data or (as we will see later) interval collapsed data
6. Expected value of at least 80% of cells was greater than or equal to 5
7. Expected value of all cells was greater than 1

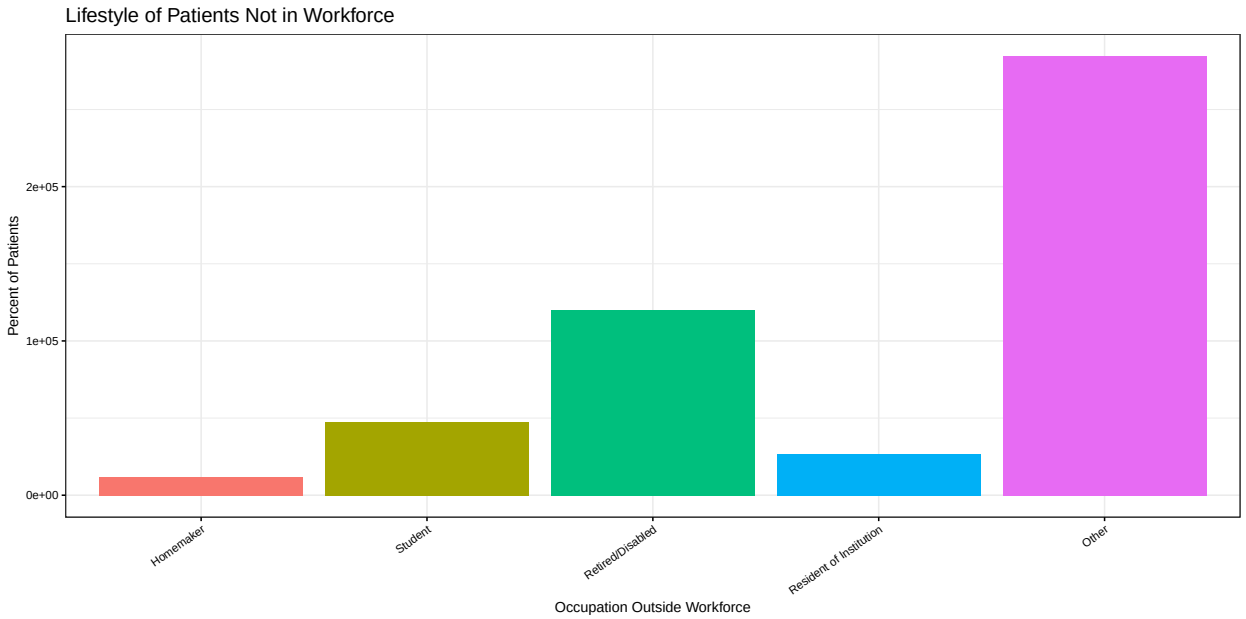
Our null hypothesis is that no association exists among the two categorical variables, and our alternative is that there is at least one association.

Correlation Plots

The Employment status by Race data satisfied all of the chi-square hypothesis test assumptions. We calculated an extremely low p-value $< 2.2e\text{-}16$, which indicated that there was a significant difference in employment by race. Our first correlation plot shows positive correlations in blue, negative in red, and the strength of the correlation is shown with size and shade. There is a really strong negative correlation between Black and Full Time & White and Not in Labor Force. There is a really strong positive correlation between Black/Other and Not in Labor Force. The second correlation plot shows the percentage each cell contributed to the test statistic that produced our p-value so close to 0. It would appear that not in labor force, especially for Black, White, and Other, were all significantly different than the other cells.

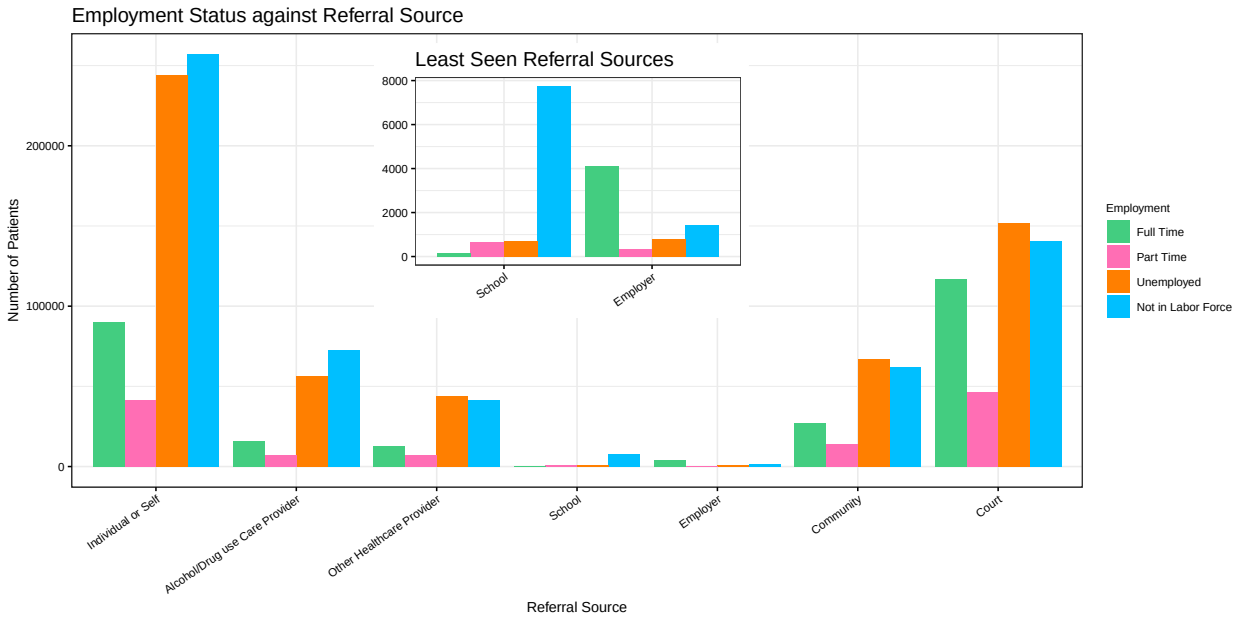
Not in Labor Force Lifestyle

Because there is a large percent of patients who fall under "not in labor force" whose further details are unknown, we decided to take a closer look at the distribution of status among this group. We can see that the largest population is other followed by retired/disabled, while homemakers are a very small percentage of this group.



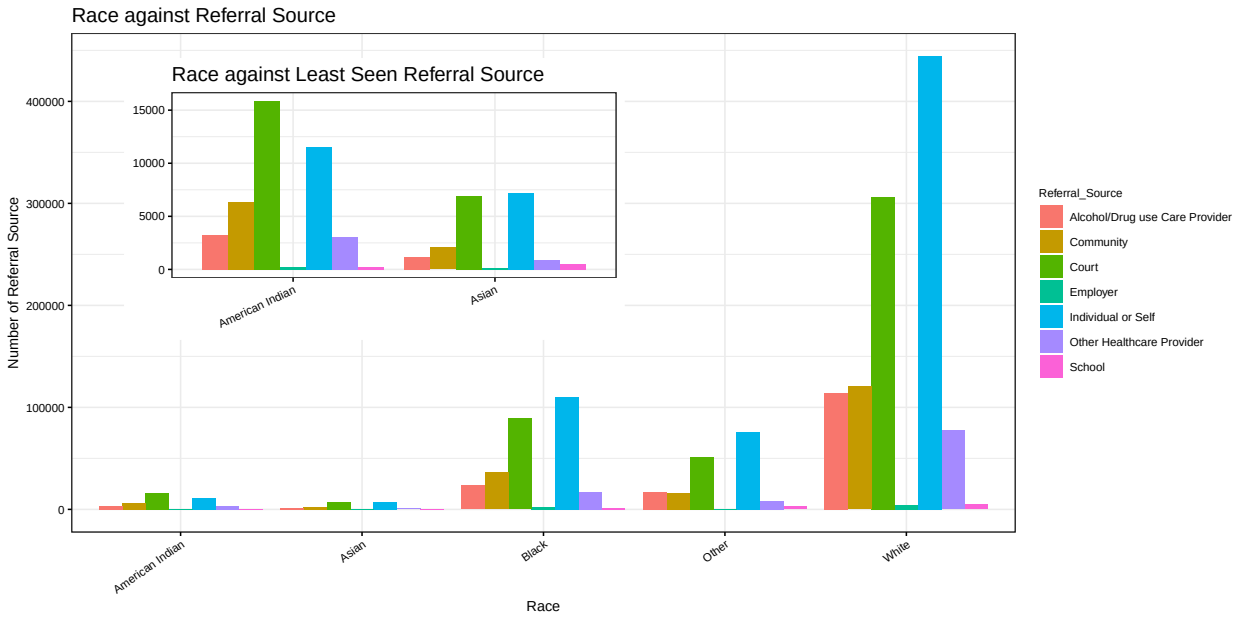
Referral Source by Employment Status

We looked at referral source by employment status to see what kind of support system the substance abusers have. Have they the presence of mind to seek help on their own? If not, who is helping them? Since most patients are unemployed/ not in labor force, it is noteworthy that most unemployed/ not in labor force patients referred themselves. Most full or part time employed patients were referred by court. When looking at the least seen referral sources, it makes sense that the highest percentage in employment status for school referrals is not in labor force as those are students, and the highest percentage in employment status for employer referrals is full time employees.



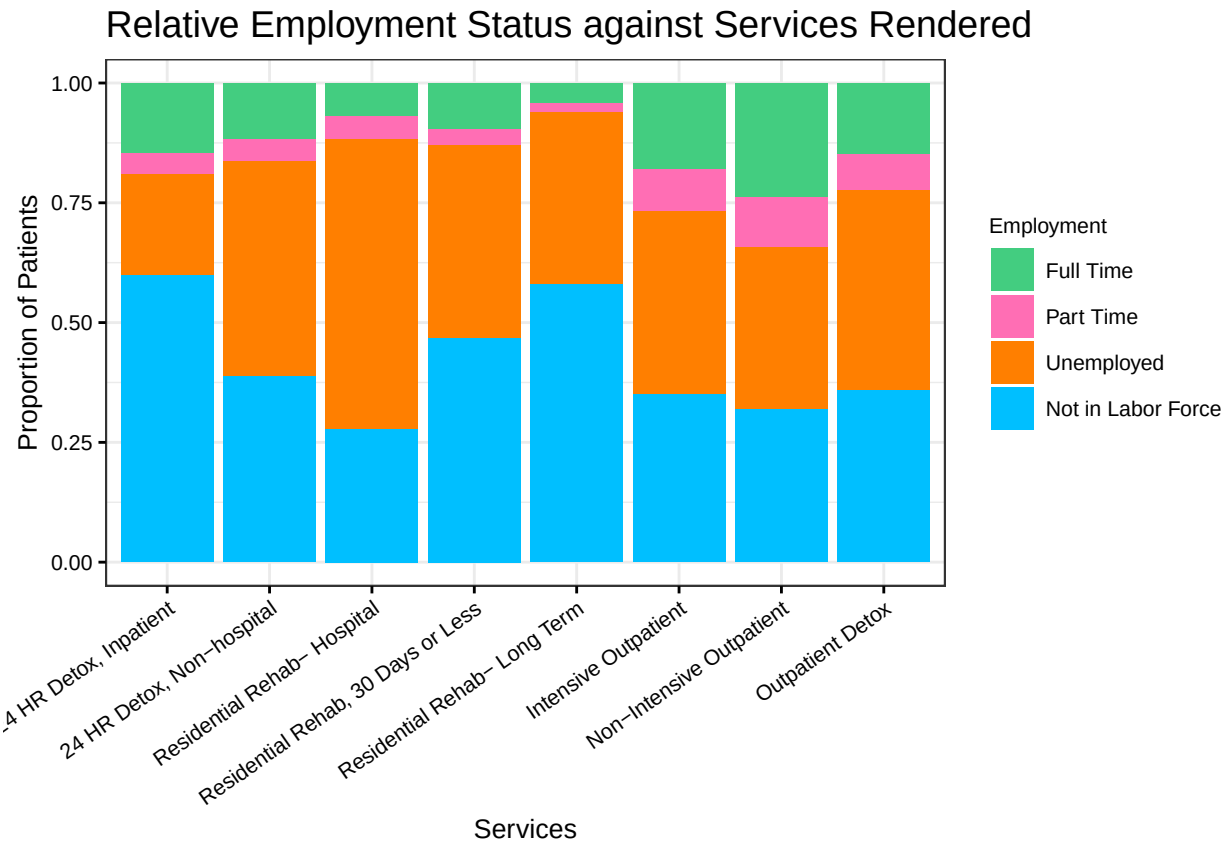
Race against Referral Source

We decided to look at referral source by race. Again we see that most patients are either self-referred or referred by court. Self-referral is higher than court-referral for White, Black, and Other races; however, court-referral is higher for American Indians. For Asians self-referral and court-referral are almost the same.



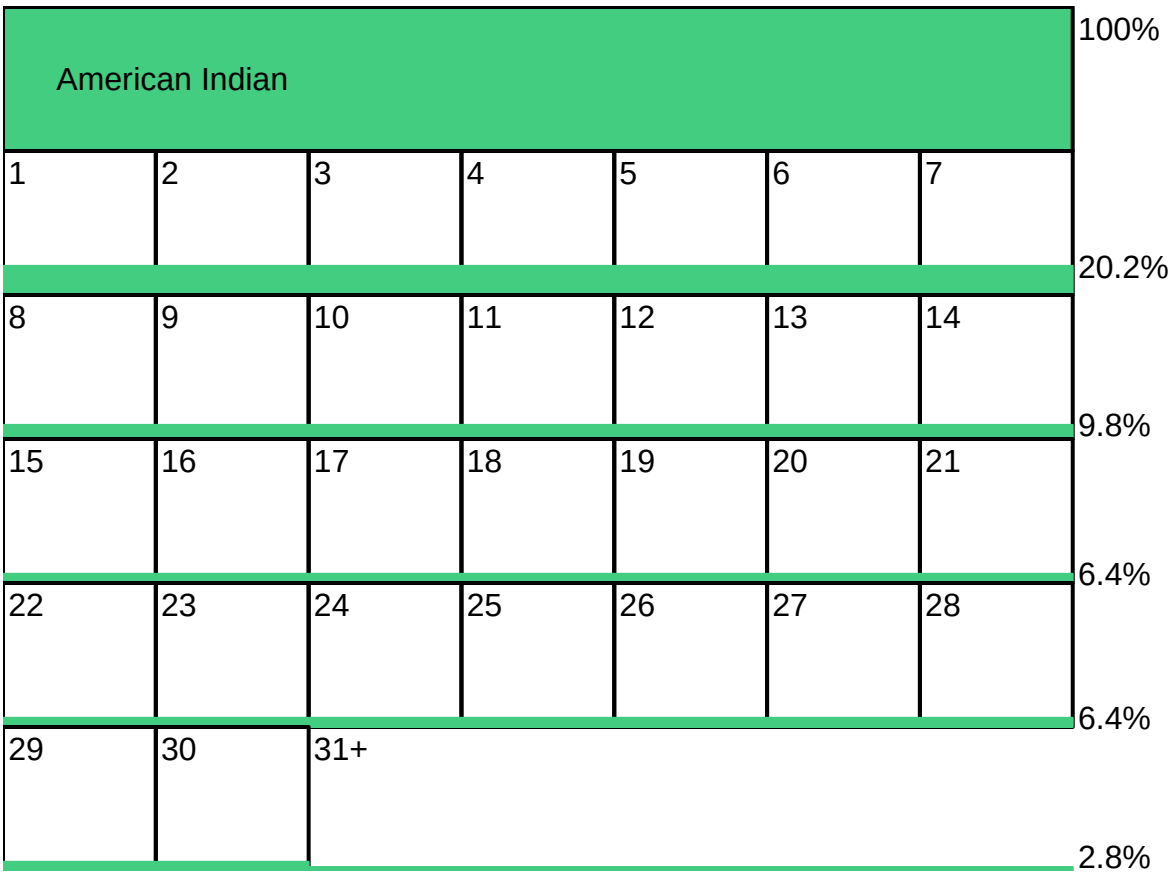
Service Category

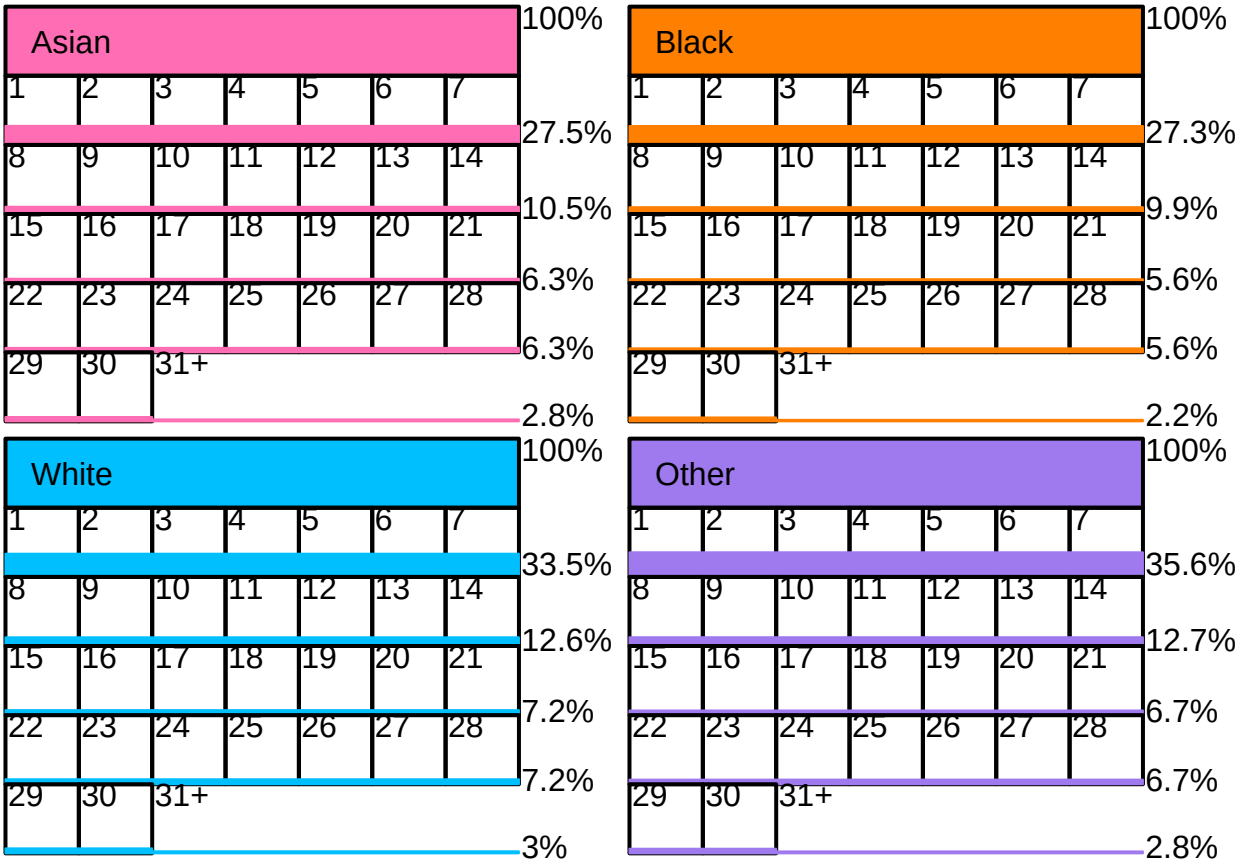
With these plots we wanted to see if there are any trends in services rendered among the patient's employment status. We want to observe how employment affects or is related to mental health and subsequent substance abuse. In the following plot we see the majority of patients for 24 hour inpatient detox and long term residential rehab were not in labor force. The majority of patients doing 24 hour detox in a non-hospital setting, residential rehab in a hospital, and outpatient detox were unemployed.



Accessibility to Treatment (Daywait)

In our own original plot, we are displaying wait time from first seeking treatment until admittance by race. We displayed this in a calendar format as the data organized the wait time by days.It is seen that absolutely everyone (100%) waited at least 0 days. 20.2% of American Indians were still waiting for at least 1-7 days, 9.8% at least 8-14 days, 6.4% 15-30 days, 2.8% 30+ days.

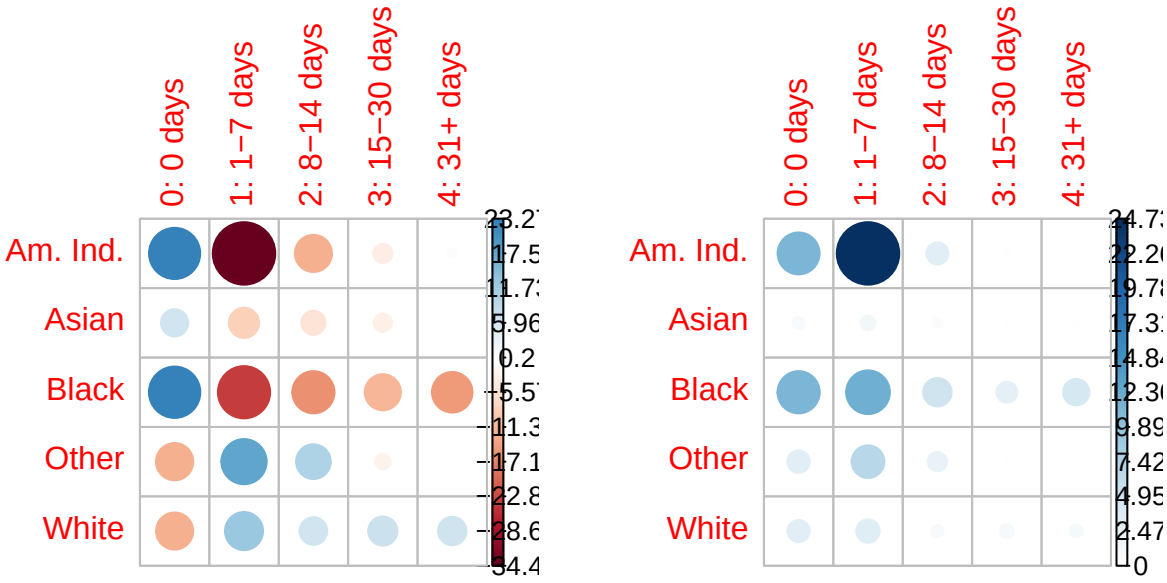

American Indian
100%
1
2
3
4
5
6
7
20.2%
8
9
10
11
12
13
14
9.8%
15
16
17
18
19
20
21
6.4%
22
23
24
25
26
27
28
6.4%
29
30
31+
2.8%



We wanted to know if there was a significant difference in wait times experienced by the different races, which would indicate a racial disparity in access to treatment. For this we used another chi-square hypothesis test. We found that at least one race had significantly different wait times.

Correlation Plots

Using the same type of chi-square test as before, we again had a very small p-value $< 2.2e\text{-}16$. This suggests that there is a difference in access to treatment by race. It is seen in the first plot that the strongest negative correlation is between the American Indian race and wait time of 1-7 days. The two strongest positively correlated cells are between the American Indian race and 0 day wait, and the black race and 0 day wait, with (as seen in the second plot) our strongest negative correlation supplying the largest percent to the test statistic.



Drug Experience

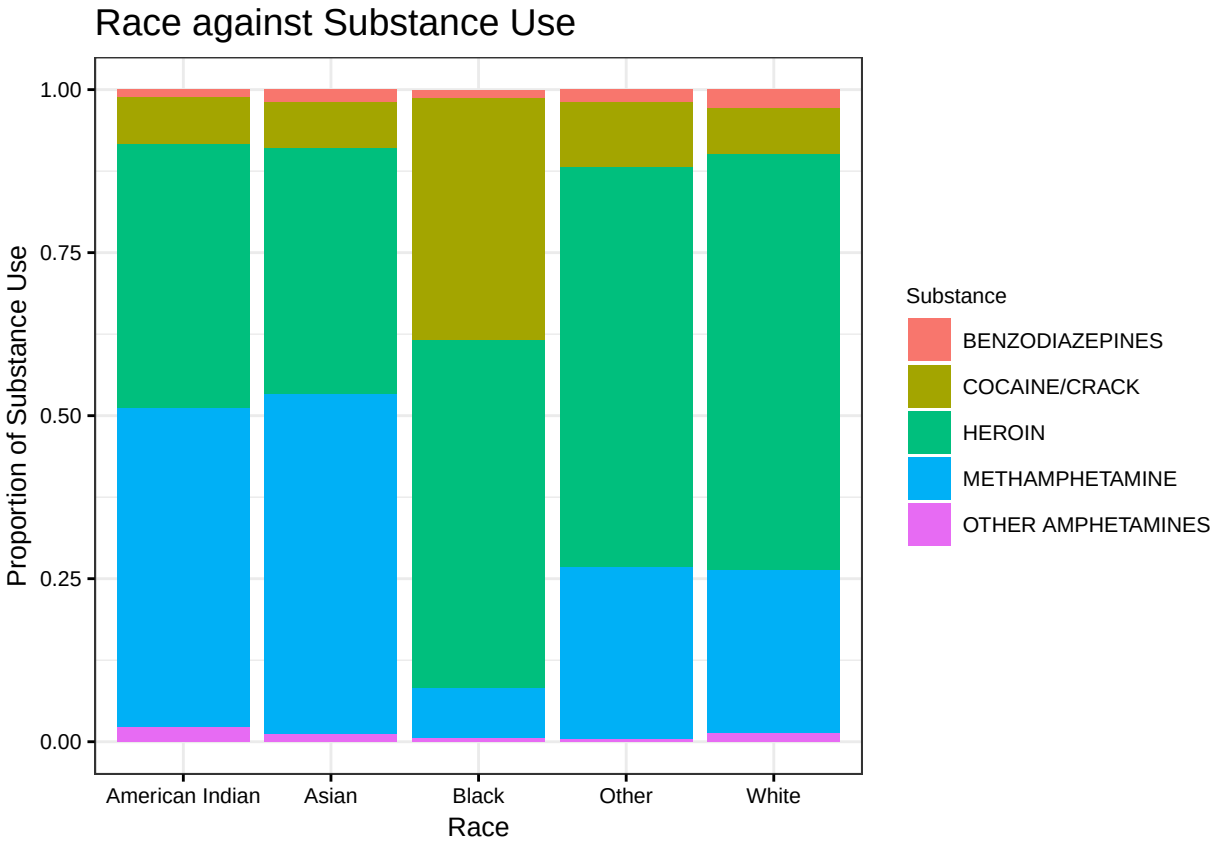
We used our tertiary data to determine if certain drugs elicited more positive or negative experiences. We chose the most frequently mentioned drugs in the scraped comments and mined the comments for positive and negative words. The most frequently seen positive and negative words in the comments were selected. The following words were chosen:
Positive: joy, happy, amazing, incredible, satisfaction
Negative: fear, scary, hate, regret, miserable

Drug	Positive Words Proportion	Negative Words Proportion
Meth	0.08	0.36
Benzodiazepines	0.17	0.21
Other Amphetamines	0.15	0.30
Cocaine	0.11	0.24
Heroin	0.26	0.41

Using our list of positive and negative words, we found that most drugs were associated with negative emotions and experiences. Meth had the highest disparity between positive and negative associations. For the other drugs, the proportion of positive associations was approximately half of the negative associations. The highest proportion of positive words was for heroin.

Admissions by Drug

From our primary data set we looked at admissions by drugs. For each race, we looked at the proportion of patients admitted for a certain drug. Most races were admitted for heroin and meth, except for black patients. Black patients were typically admitted for cocaine and heroin. It is noteworthy that the highest proportion of positive experiences from the tertiary data set was in heroin and the highest percentage of admissions by drug across most races was heroin.



Conclusion

White and Other race substance abusers seemed to suffer from depression the longest. The age that black patients are admitted does not taper down in older years, but instead seems bi-modal with a second peak at ages 50-54. Most patients are unemployed/ not in labor force. Self-referral and court-referral were most common across all races. The strongest negative correlation for race and wait time is between the American Indian race and wait time of 1-7 days. Most races were admitted for heroin and meth; furthermore, heroin had the highest proportion of positive experiences of all drugs mentioned from the tertiary data set.