Data Visualization

George Rhodes

12/6/2017

The following visualizations were created using data from the National Survey on Drug Use and Health from 2015 (public use data set), which I am using to explore the relationship between socioeconomic status (SES) and recovery from addiction. I hypothesized that those with higher SES would have higher rates of recovery due to fewer psychosocial stressors and better access to resources. While I have yet to show strong statistical support for this hypothesis, these exploratory visualizations do reveal interesting trends.

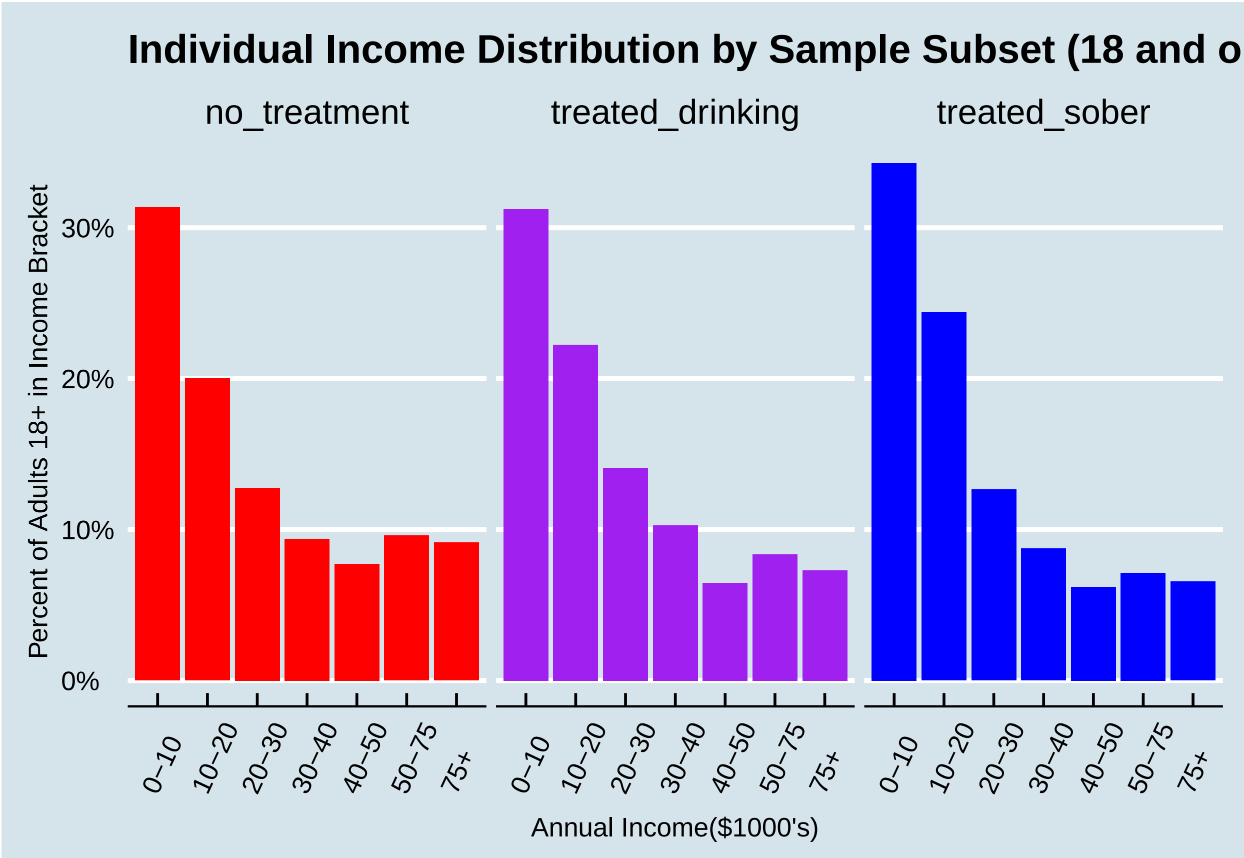
The operationalization of these visualizations help to convey the socioeconomic landscape of the sample and those that have been treated for addiction. I created a new catagorical variable to mark those who have not been treated for addiction (no_treatment), those who have been treated and are still drinking (treated_drinking), and those who have been treated and have not drank in the last year (treated_sober). By looking at these three seperate sample subsets side by side with faceted geom_bars, we can begin to see how they vary by SES.

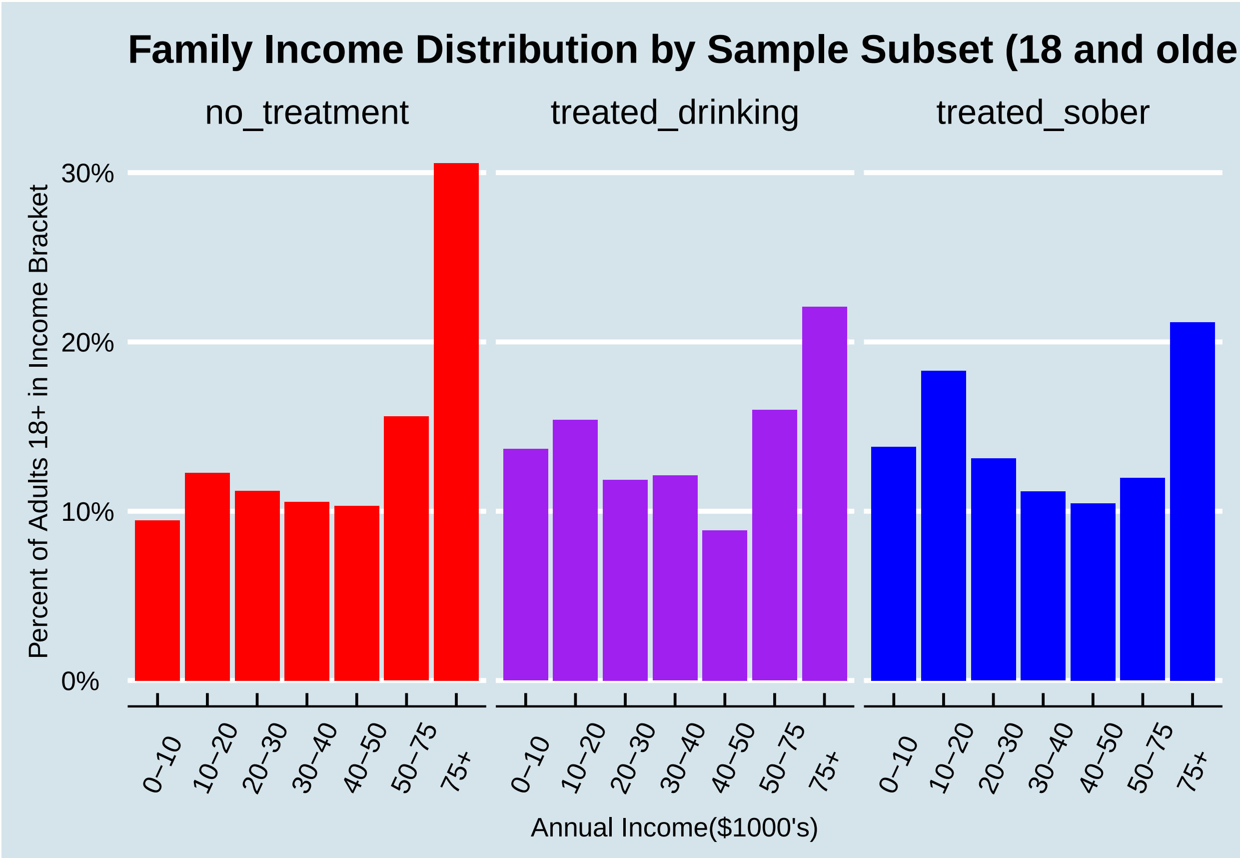
The three visualizations are essentially set up the same way. Only the x-axis varies between the three to look at income, family income, and level of educaiotn, in an attempt to measure of SES. The Y-axis is the percent of the sample subset–either no_treatment, treated_sober, or treated_drinking depending on the facet. So each bar shows the percent of the sample subset that has each level of SES measure. While this may seem difficult to understand intially, it is nececcary to compare the within-group profiles because the vase majority of the survey's sample has never recieved treatment. Therefore it is more helpful to compare these proportions with subsets than of the entire sample. showing these proportions graphically was another challenge. It took a while to figure out that I needed a seperate geom_bar layer for each population so that the proportion was made with the appropriate count.

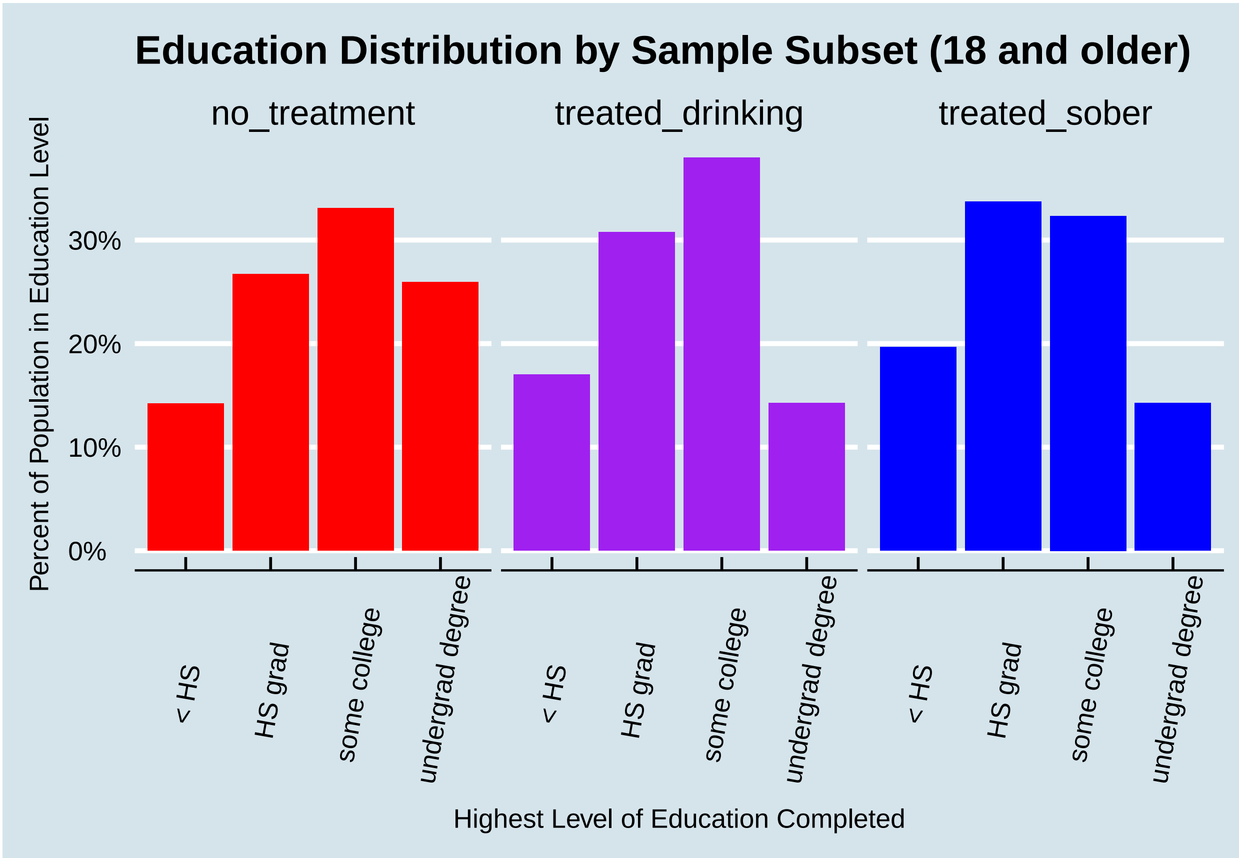
I did have to manipule the data a bit. I filted out respondants under 18 to avoid confusion with the education level and becasue they were disproportionately on the low end of income measures. I also had to change the survey responses from integars to strings so that graphs could display the labels of each catagory. Then I had to play with the angle and alignment along the x-axis for easier interpretation. From there it was mostly just aesthetic tweaking.

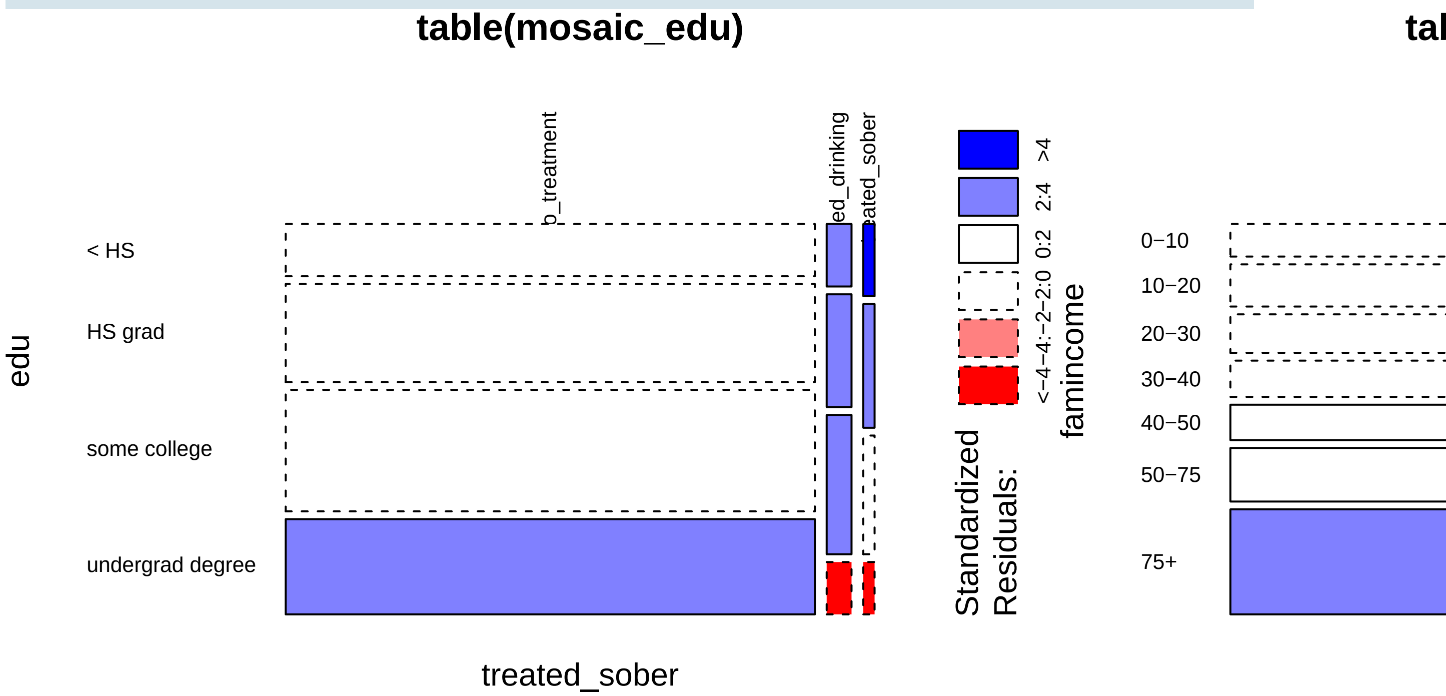
While my data does not show the associations I was expecting, I chose these visualizations because the findings are interesting. For the most part, those who have been to treatment are of lower SES, while the group that has not been to treatment has much larger proportions in the highest brackets of eadh catagory. The could be becasue those of lower SES have higher addiction rates, or it could mean that those of higher SES are less likely to have been treated. Both explanations fit the theory. These visualizations are helful for understanding treated and untreated populations yet clearly require further exploration.

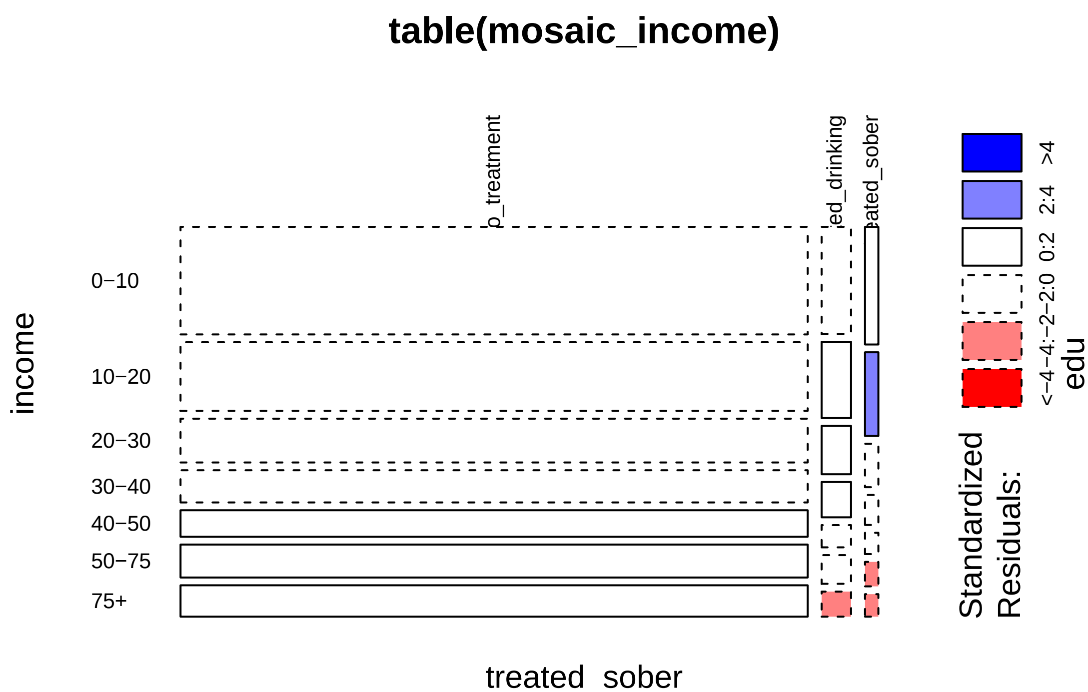
```
## Warning: package 'dplyr' was built under R version 3.4.2
```

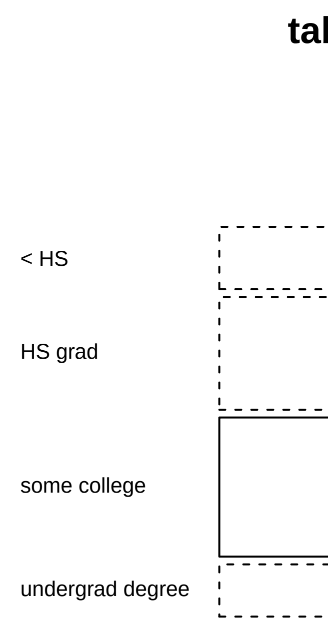

Individual Income Distribution by Sample Subset (18 and ol
no_treatment
treated_drinking
treated_sober
Percent of Adults 18+ in Income Bracket
30%
20%
10%
0%
0–10
10–20
20–30
30–40
40–50
50–75
75+
Annual Income($1000's)

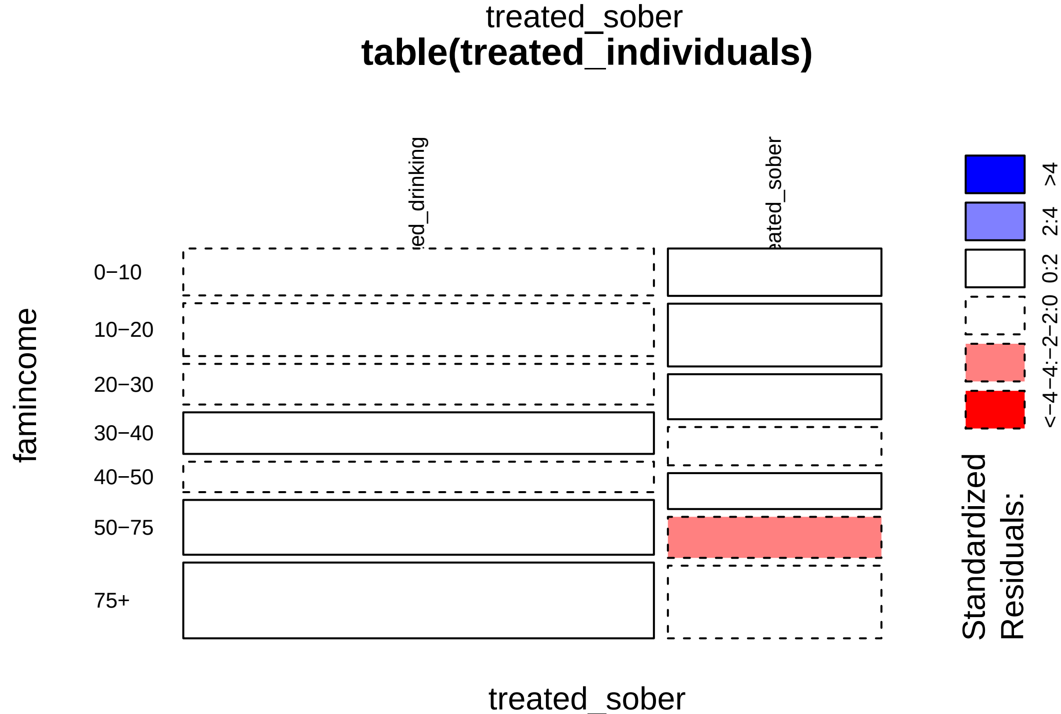

Family Income Distribution by Sample Subset (18 and older
no_treatment
treated_drinking
treated_sober
Percent of Adults 18+ in Income Bracket
30%
20%
10%
0%
0–10
10–20
20–30
30–40
40–50
50–75
75+
Annual Income($1000's)


Education Distribution by Sample Subset (18 and older)
no_treatment
treated_drinking
treated_sober
Percent of Population in Education Level
0%
10%
20%
30%
< HS
HS grad
some college
undergrad degree
Highest Level of Education Completed


table(mosaic_edu)
edu
< HS
HS grad
some college
undergrad degree
treated_sober
Standardized Residuals:
>4
2:4
0:2
-2:0
-4:-2
<-4
famincome
0–10
10–20
20–30
30–40
40–50
50–75
75+


table(mosaic_income)
p_treatment
ed_drinking
ated_sober
income
0−10
10−20
20−30
30−40
40−50
50−75
75+
treated_sober
Standardized Residuals:
<−4
−4:−2
−2:0
0:2
2:4
>4


tal
edu
< HS
HS grad
some college
undergrad degree


table(treated_individuals)
ed_drinking
eated_sober
famincome
0−10
10−20
20−30
30−40
40−50
50−75
75+
treated_sober
Standardized Residuals:
<−4
−4:−2
−2:0
0:2
2:4
>4