Lab 11 - Data, Aesthetics, & Geometries

George Rhodes

November 9, 2017

Complete the following exercises below. Knit together the PDF document and commit both the Lab 11 RMD file and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Which variables in your dataset are you interested in visualizing? Describe the level of measurement of these variables and what type of geography you think is appropriate to represent these variables. Give your reasoning for choosing the `geom_()` you selected.

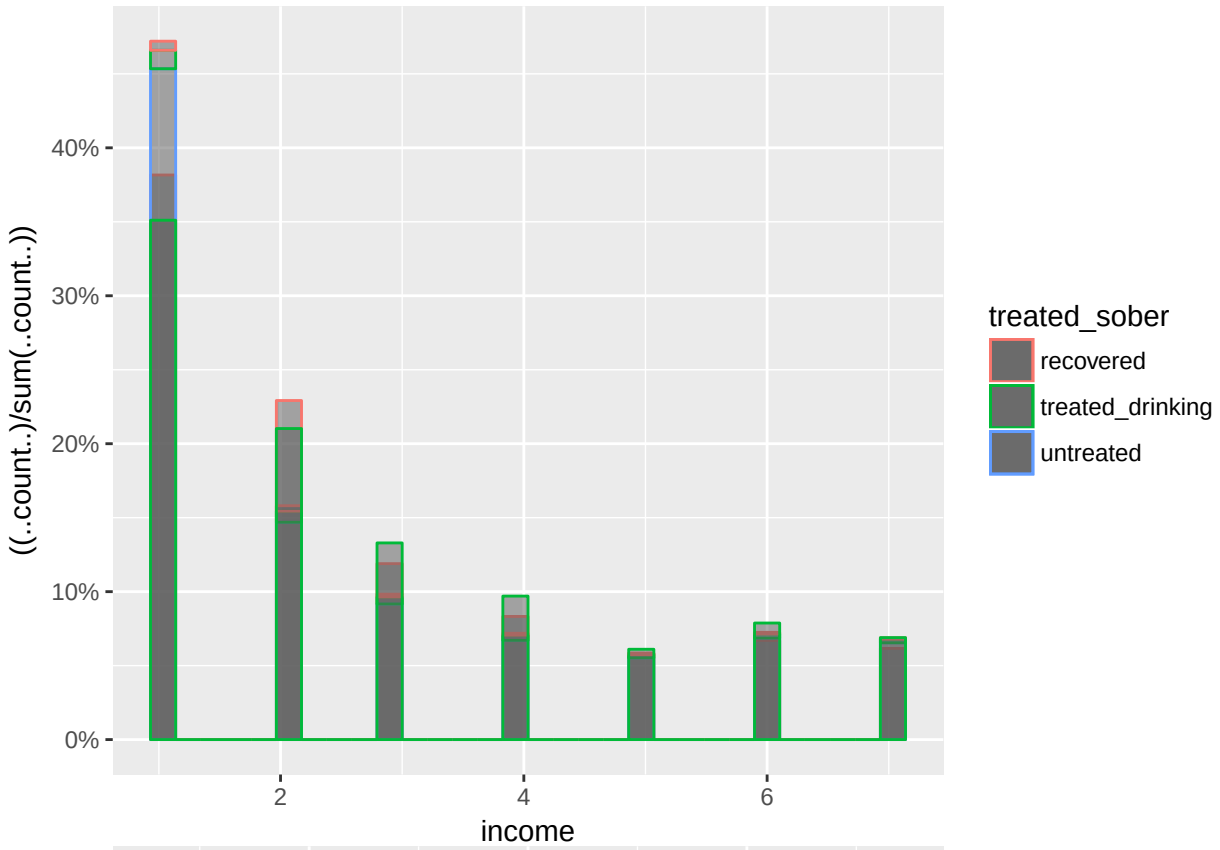
The countytype and poverty variables are nominal. Showing them as percentages would be helpful. compare to general popluation? Income, fam income, and edu as histograms becasue they are ordinal. Compare with a layer.

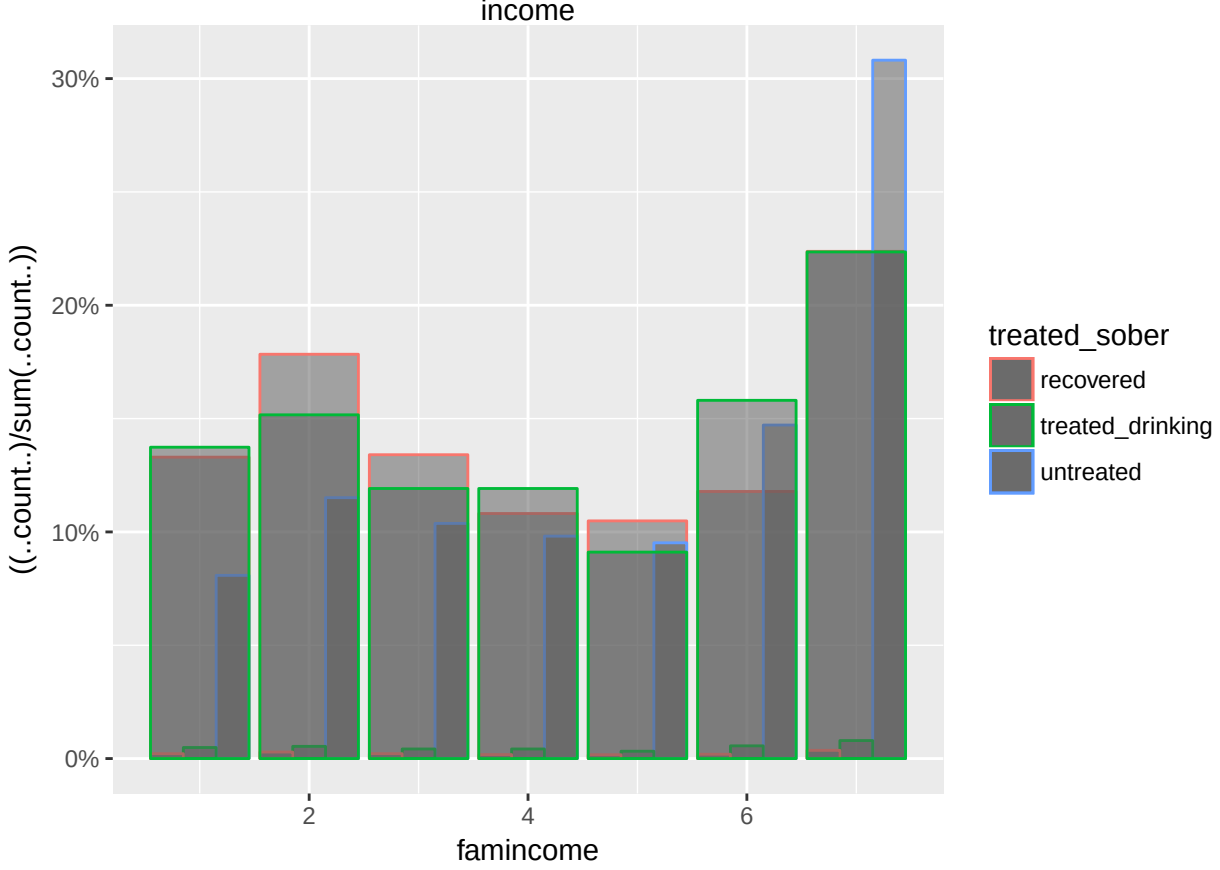
The layer to compare with could come from the overal survey data of total respondants, or compare to those that have received treatment but have not been sober. This second idea might require additional variables of each drug, which would make the analysis more robust overall. Need to work this part out.

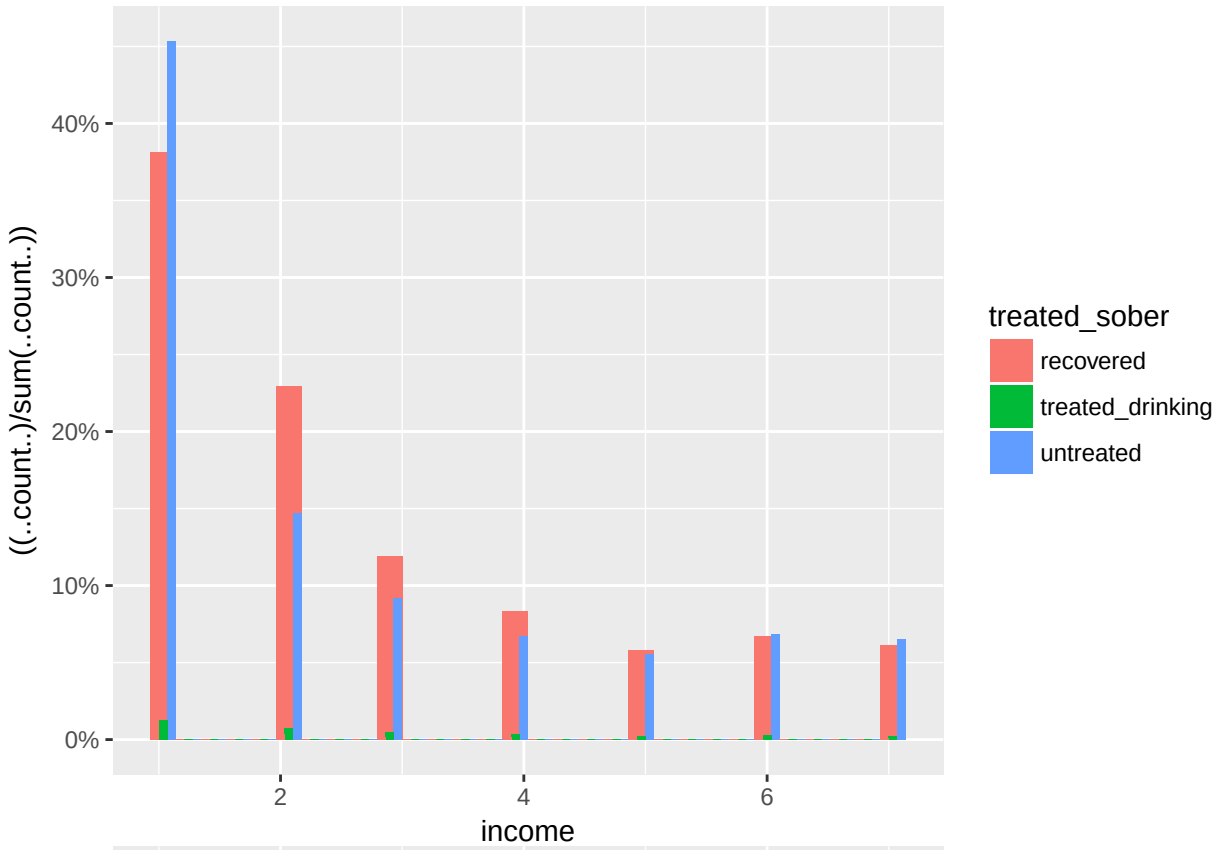
```
## Warning: package 'dplyr' was built under R version 3.4.2
```

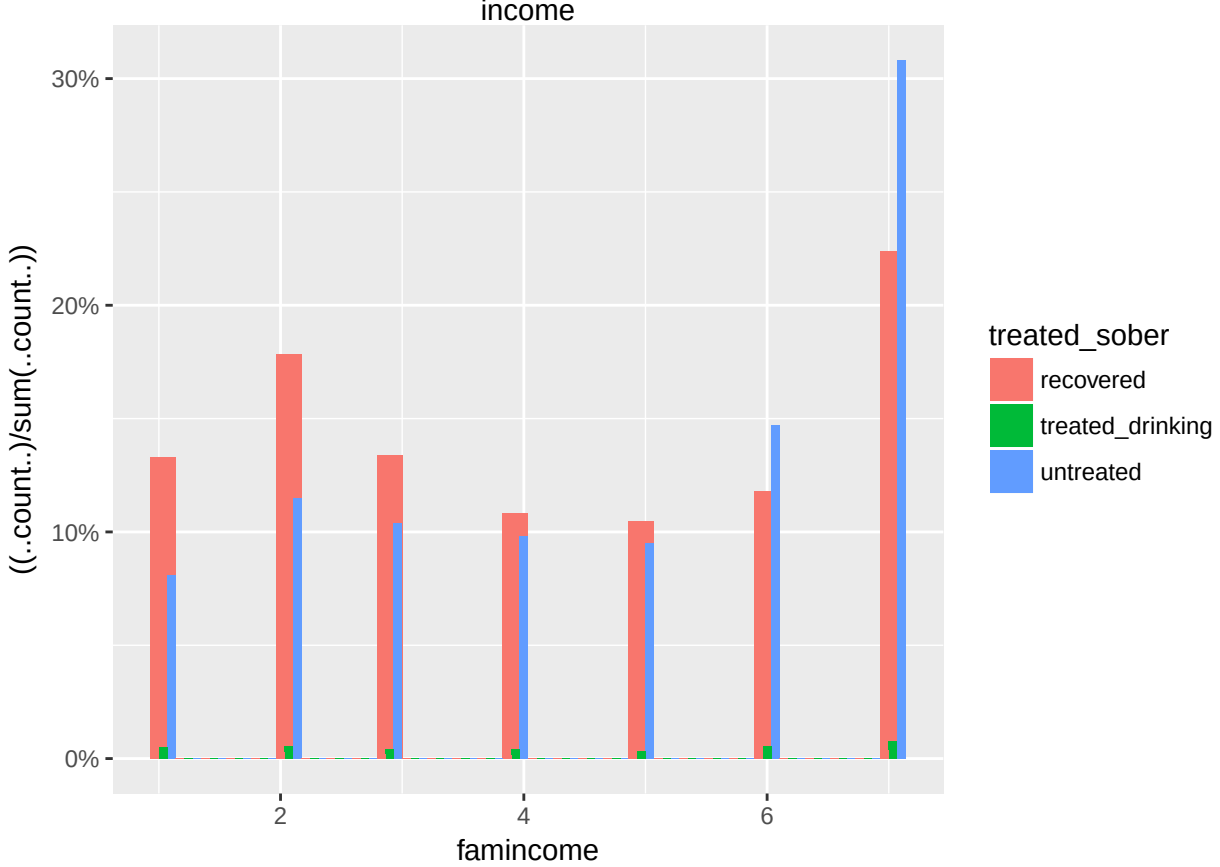
2. Is your data in the proper format to visualize the data in the way you want? Why or why not? *If you need/want to change the structure of your data, do it below.*

3. Create at least two different exploratory plots of the variables you chose using the skills we covered in class today. What types of mapping aesthetics did you choose and why? What do these plots tell you about your data?


40%
30%
20%
10%
0%
((..count..)/sum(..count..))
2
4
6
income
treated_sober
recovered
treated_drinking
untreated


30%
20%
10%
0%
((..count..)/sum(..count..))
2
4
6
famincome
treated_sober
recovered
treated_drinking
untreated





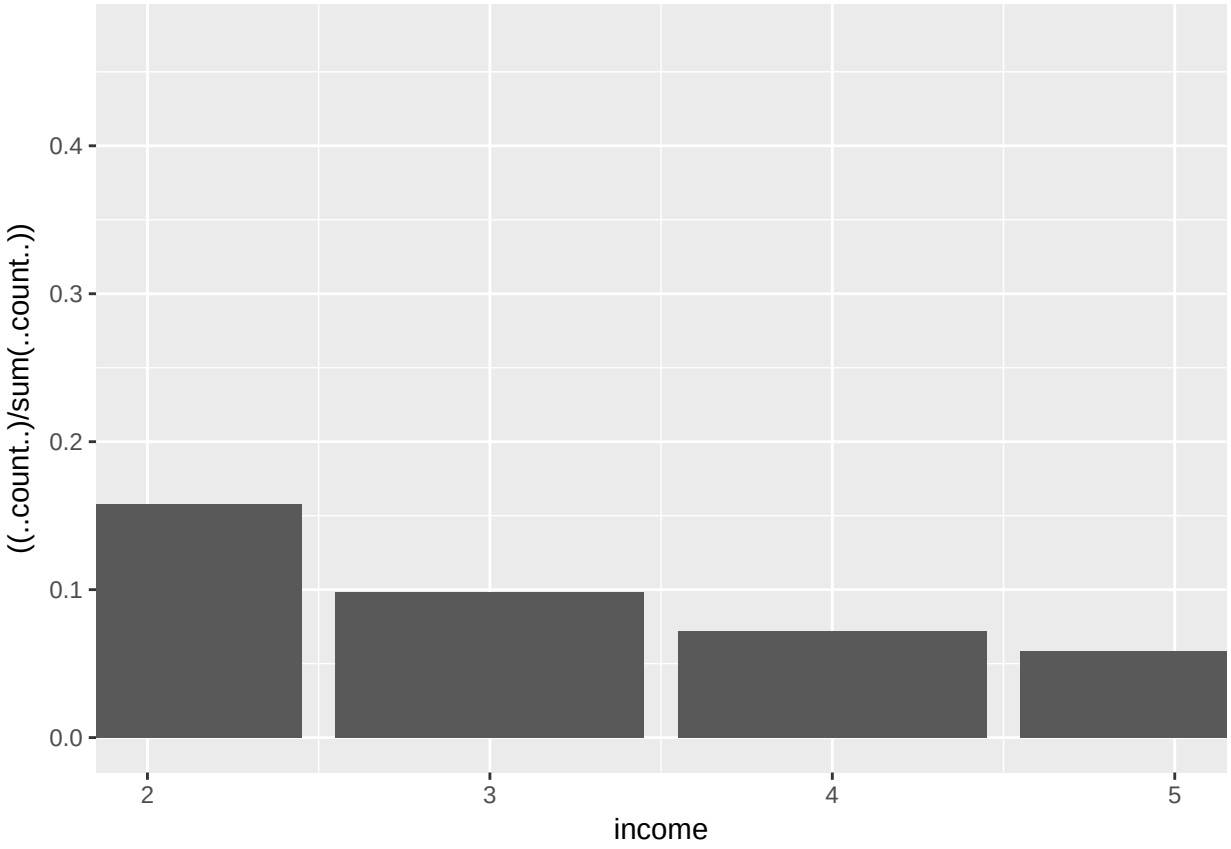
#geom_bar(recovered_respondants, aes(y = ((..count..)/sum(..count)), x = income_variables_recovered))

I want to show the edu of respondants by whether they are treated or untreated under the column "treated_sober" on the hist_edu bar char.

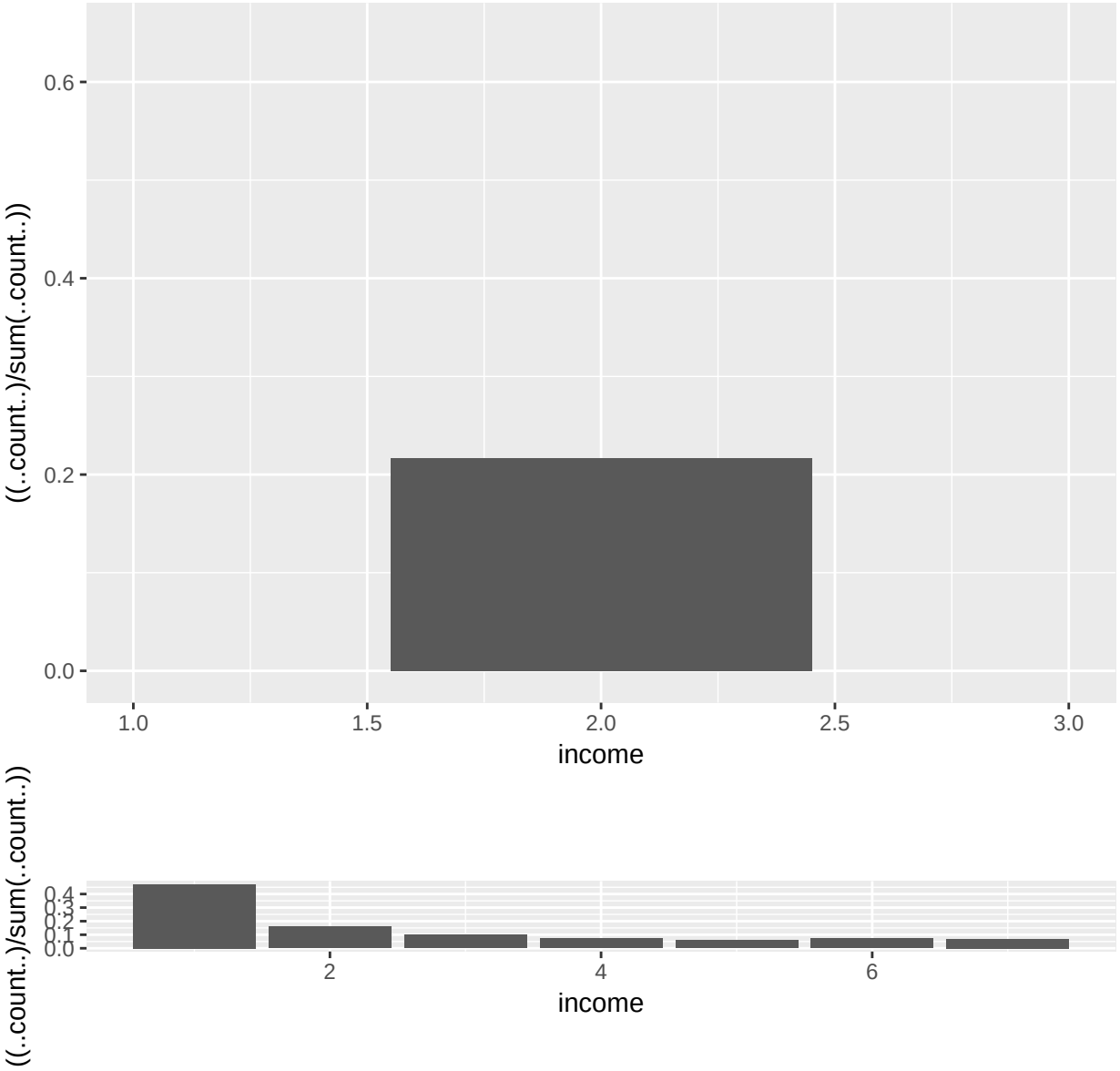
I also want to to do the same thing for hist_income bar chart. This is losing me.

I think I need to add some more aesthetics: label the y-axis as percent. Change labes of x-axis from 1-5 to the correstponding income/edu values of the codebook. I'm not sure where or how to do this. Add the layers so I can compare the edu profile of those that are recovered from those that are unrecovered (the rest of the population) Same additional later for income.

4. Create at least three variations of the plots you've already made by modifying some of the arguments we covered in class (i.e. `position`, `scale`, `size`, `linetype` etc.). Do any of these modifications help you understand your data better? Why or why not? Do any of them create a misleading interpretation of the relationships between your variables? If yes, how so?




```
## Warning: Removed 15519 rows containing non-finite values (stat_count).
```



5.

From the plots you've created thus far, do any of them seem appropriate for a general audience? Why or why not? If so, what do you think you'd still need to do to make them more suitable as explanatory visualizations?

I still need to change the x axis labels to what they represent. Should I do this in the plot, or should I go back and change the values in the data?

I need to change the label of the y-axis

I need to compare the profiles of those who have recovered from treatement, with those who have not, or with the general population. I need to get all the data on the same visualizations and I am struggling.

UPDATE 12/3/17: I have worked through a lot of issues since this lab, but this is the foundation of my visualizations.