Loan EDLD 651 Lab 3

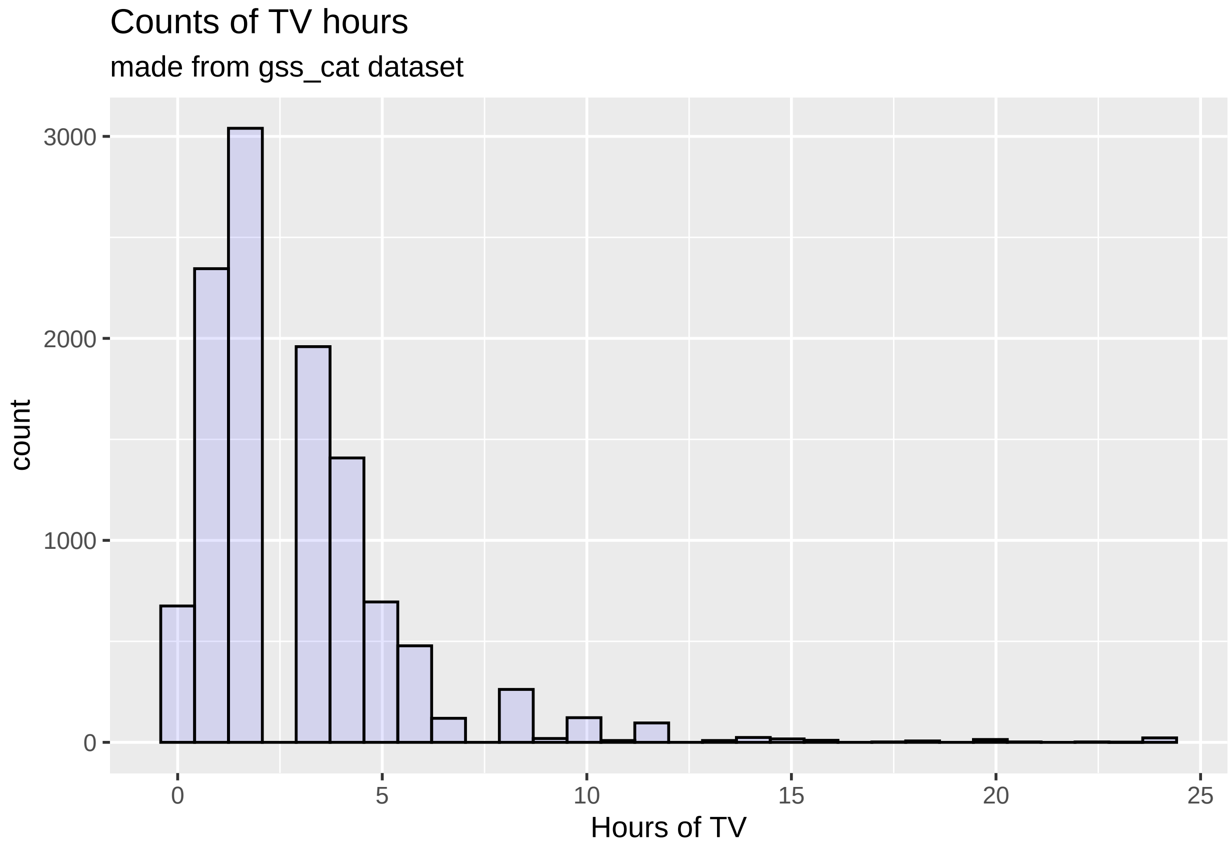
Chris Loan

1. Run the following code to load the *tidyverse*.
2. Run the following lines of code to view and read about the data we'll use for this lab.

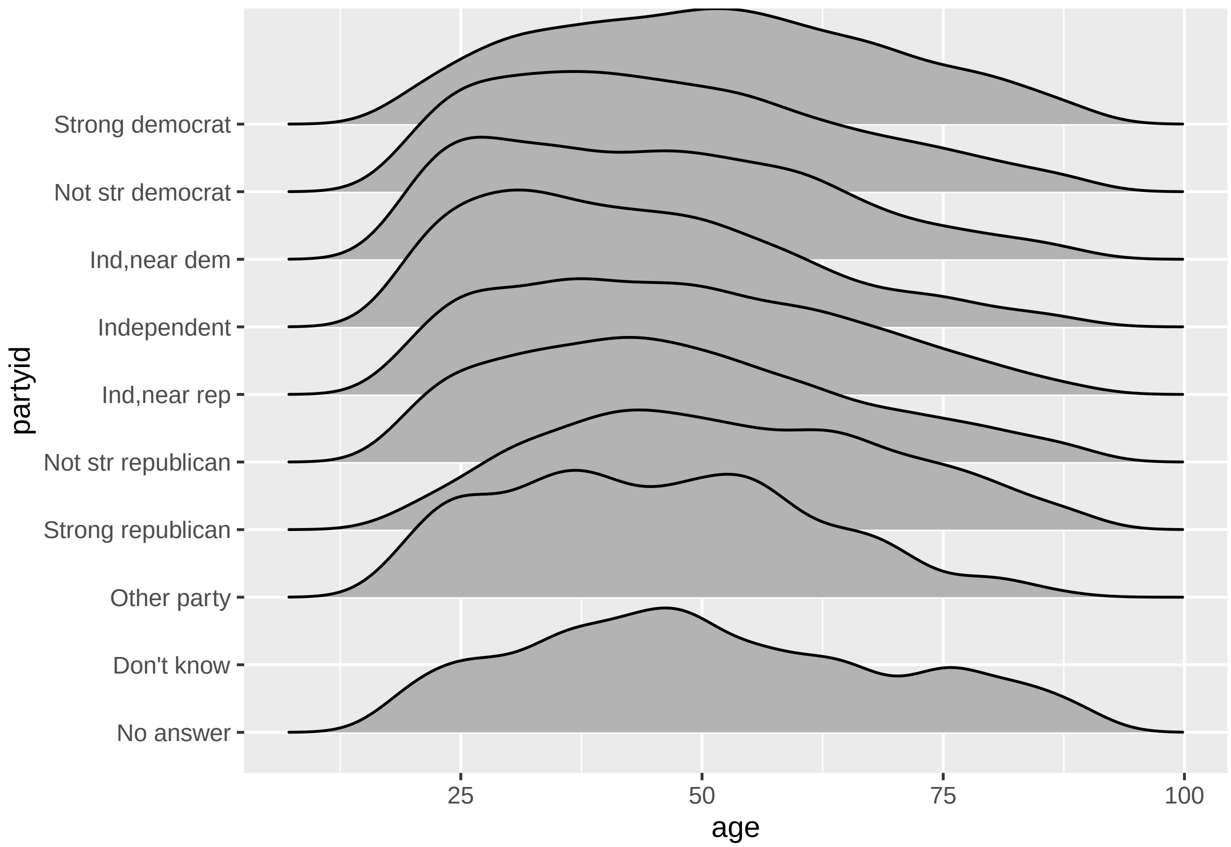
Comment out line 26 after you run it, but puting a `#` at the start.

2. Use `ggplot` to create a histogram of `tvhours`. The plot should:

- Include a small amount of transparency so gridlines are visible under the histogram.
- Be a different color than the default
- Include meaningful axis labels
- Have a title and a subtitle

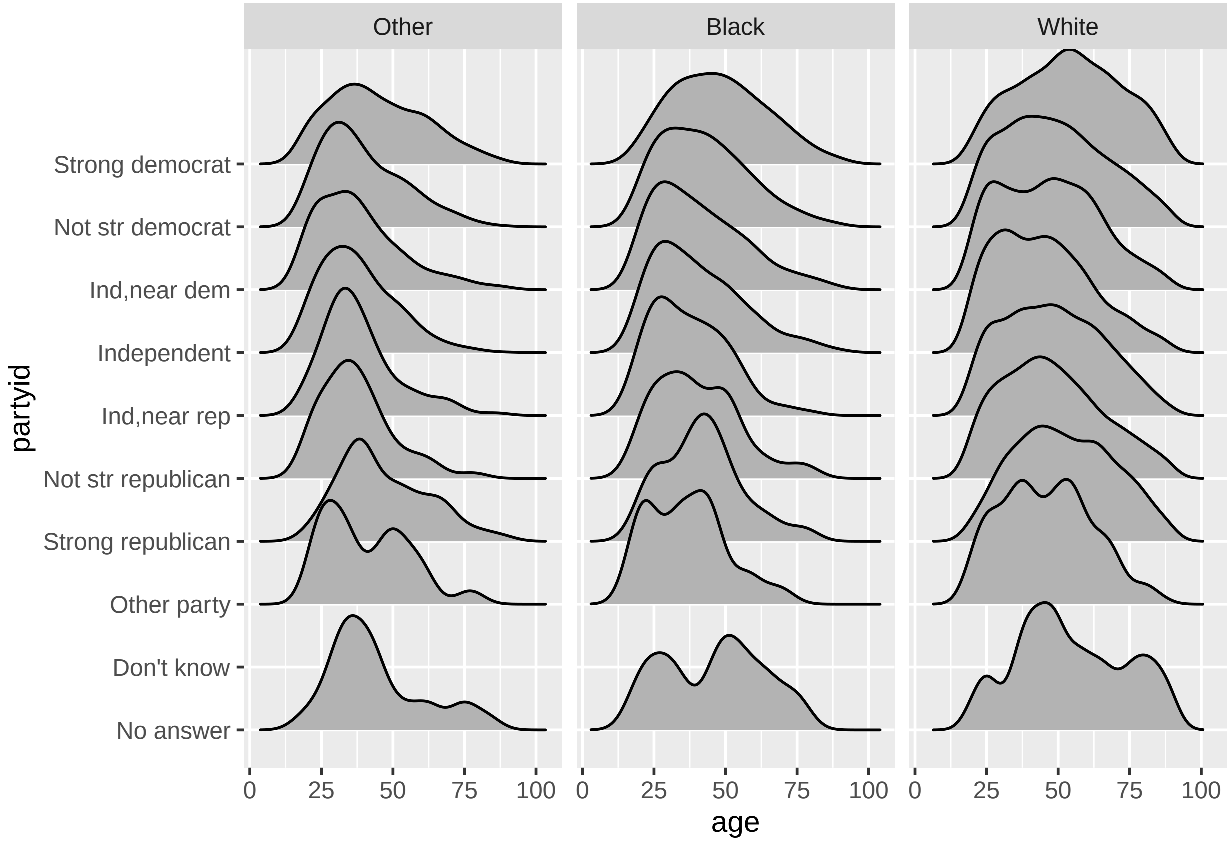


3. Create a ridgeline density plot of `age` by `partyid`. Do you notice any strong differences by party affiliation?



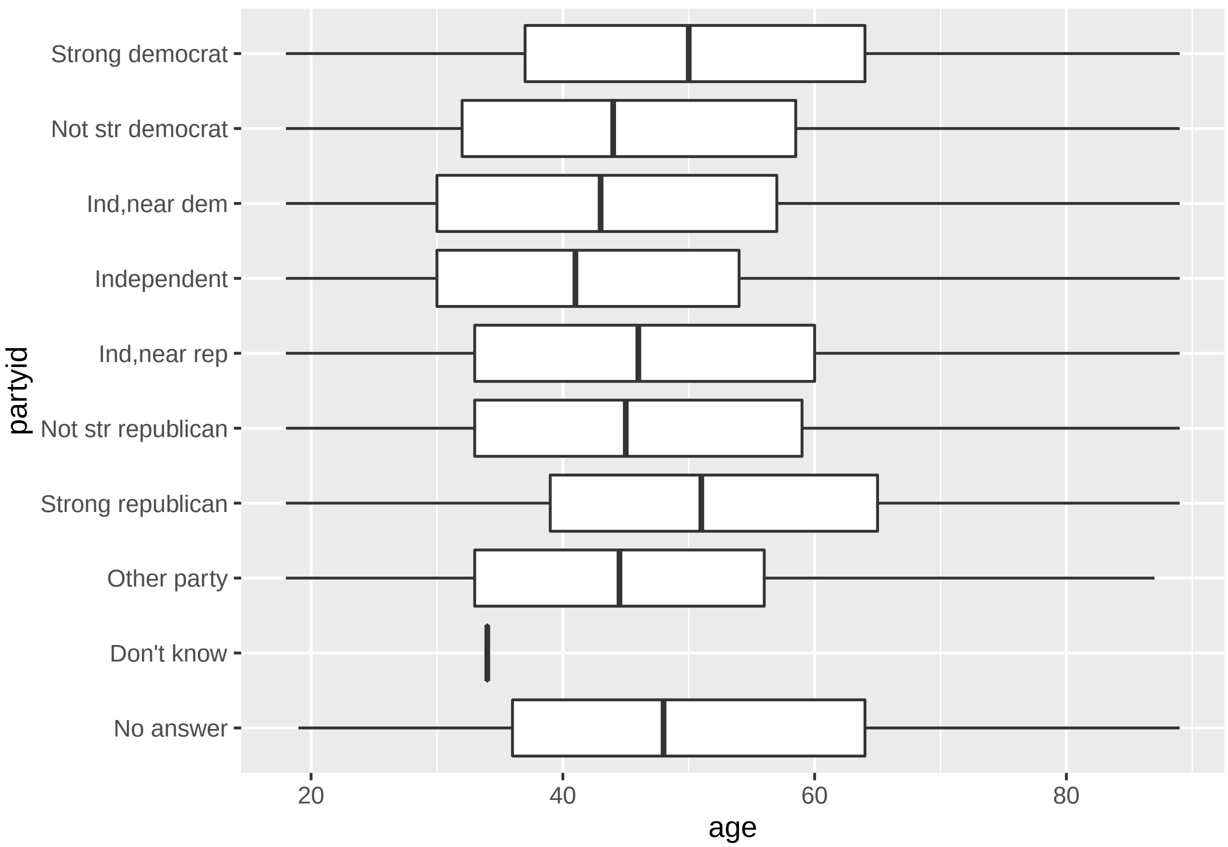
```
* With only visual inspection, it's difficult to tell (and I feel I'm not good at density ridges yet; i
* Though, it appears that older people are more likely to endorse strong affiliation with a party, compa
```

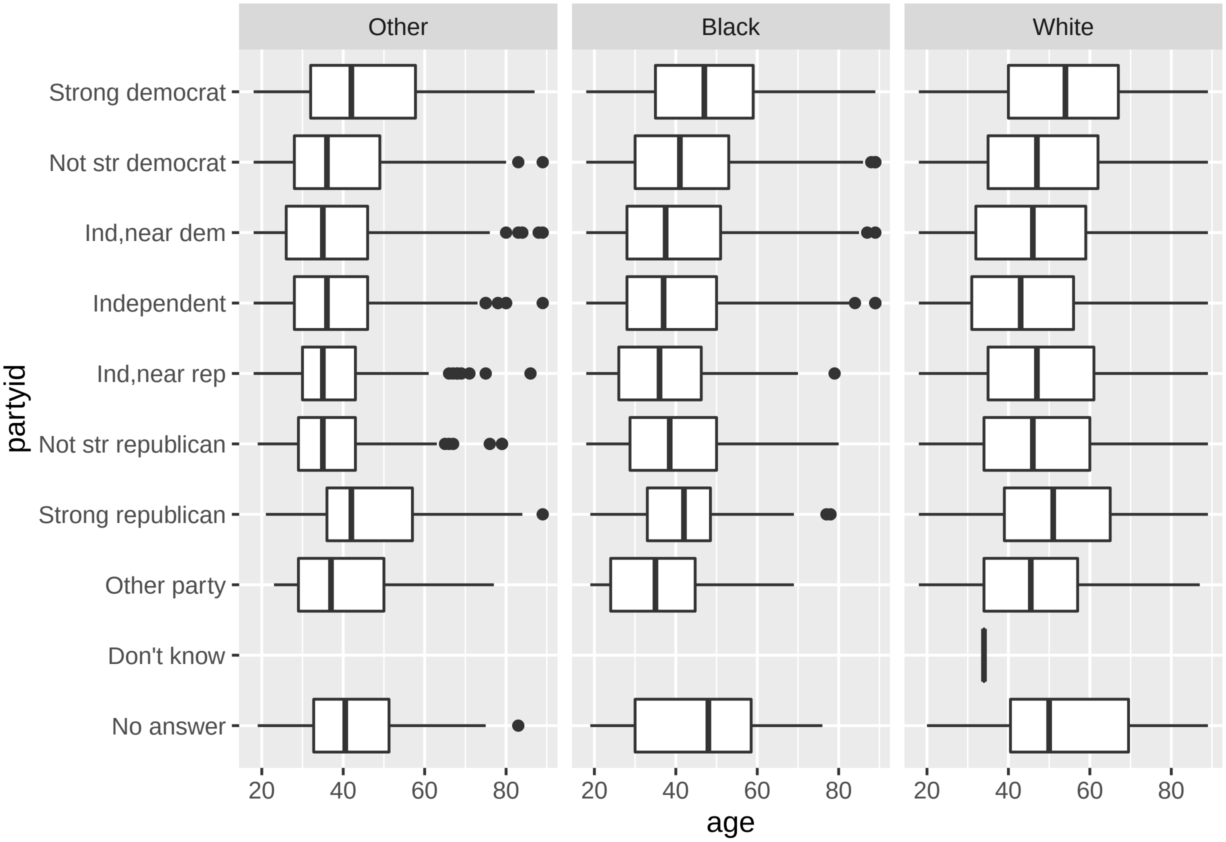
4. Modify the plot from Question 3 to produce separate ridgeline densities by `race`. Do you see any patterns?



```
* To me, I still see the same pattern, with more partisan endorsement as age increases.
* This trend appears similar across age, but again, I'm not certain it is significant.
```

5. Replicate the plots from Questions 3 and 4, but this time displaying the relations with boxplots (i.e., one boxplot from Question 3 and one boxplot for Question 4 in the chunk below). Which do you think displays the relation more clearly? Why?

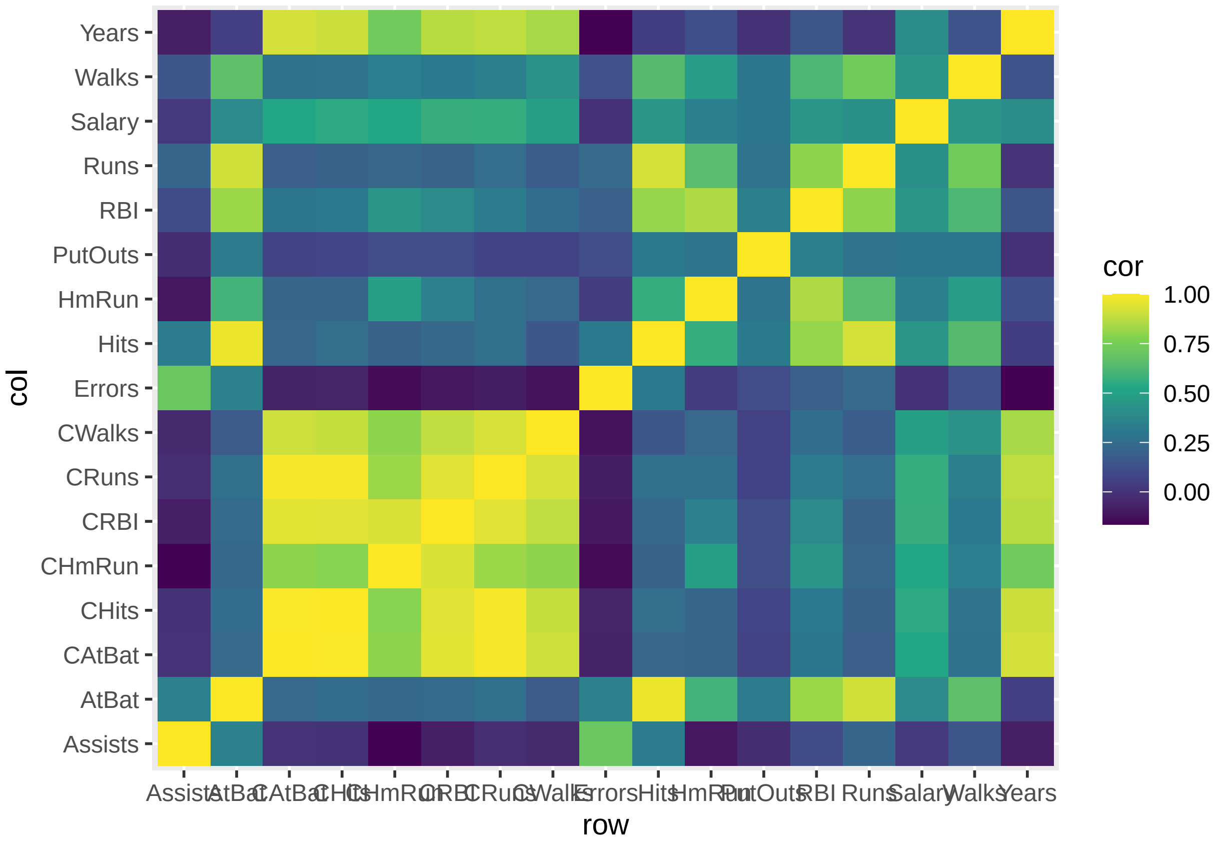

Strong democrat
Not str democrat
Ind,near dem
Independent
Ind,near rep
Not str republican
Strong republican
Other party
Don't know
No answer
partyid
20
40
60
80
age



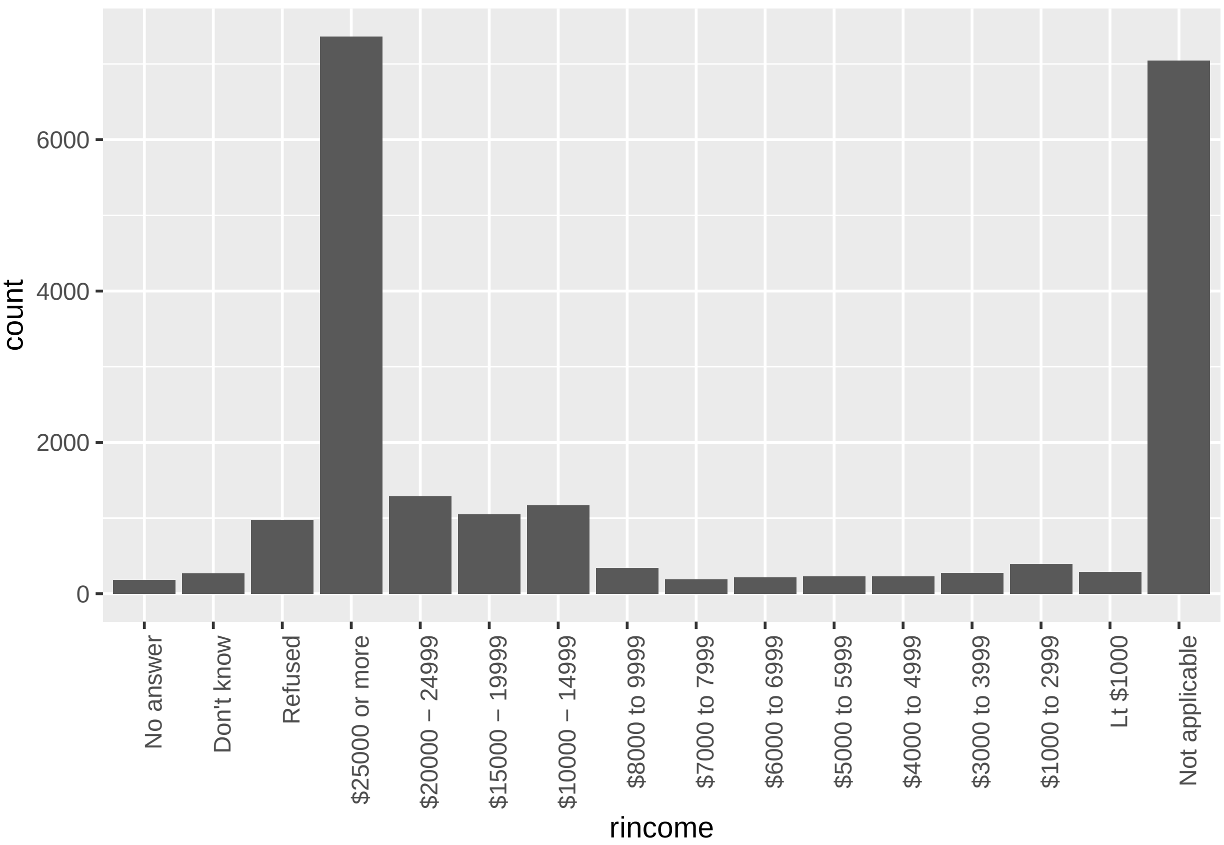
```
* I personally prefer the boxplots, but that's mainly because I'm more confident in interpretting them.
* I think the density plots are nice for showing a distribution is normal (or follows another distribut
```

6. Run the following lines of code to get a correlation table as a data frame.

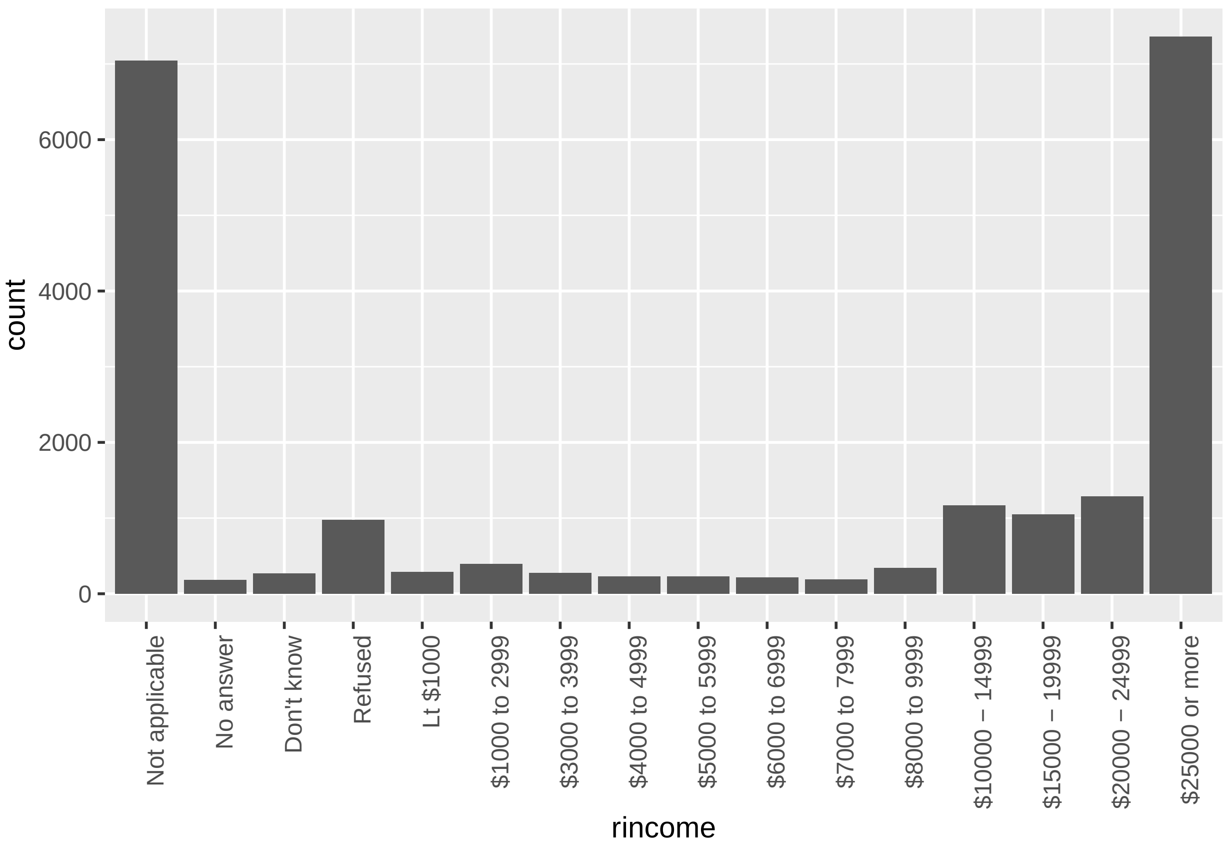
Use the correlation table to create a heatmap using `geom_tile`. Change the colors to one of the four viridis palettes. What does the heatmap help you see?



7. The code below will produce a barplot.



Use the code below to create a new data set - `gss_cat_new` - that redefines the `rincome` variable so that *Not applicable* is displayed first, followed by *No answer*, *Don't know*, *Refused*, and then each income category from lowest to highest. Once you've redefined the factor variable, re-produce the plot above by running the `ggplot` code at the bottom of the chunk. The plot should have *Not applicable* closest to the y-axis, then *No answer*, *Don't know*, etc.


count
6000
4000
2000
0
Not applicable
No answer
Don't know
Refused
Lt $1000
$1000 to 2999
$3000 to 3999
$4000 to 4999
$5000 to 5999
$6000 to 6999
$7000 to 7999
$8000 to 9999
$10000 – 14999
$15000 – 19999
$20000 – 24999
$25000 or more
rincome