Lab 1 - Data visualization

Sydney Girvin

Load Packages

```
library(tidyverse)
```

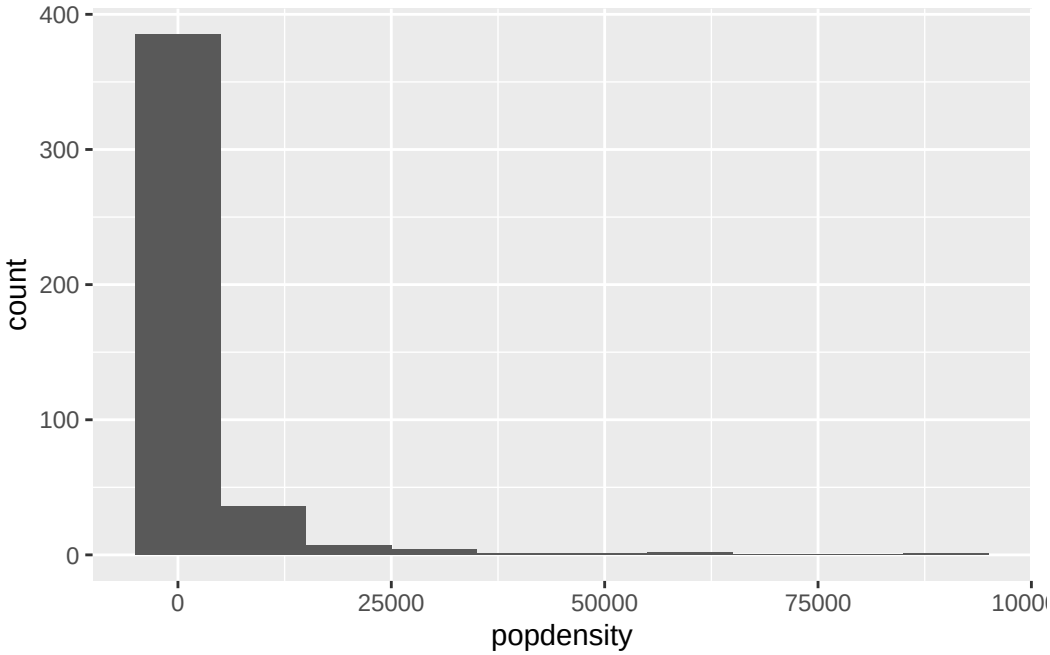
```
Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
had status 1
```

```
library(viridis)
```

Exercise 1

(Type your answer to Exercise 1 here. Add code chunks as needed. Don't forget to label your code chunk. Do not use spaces in code chunk labels.)

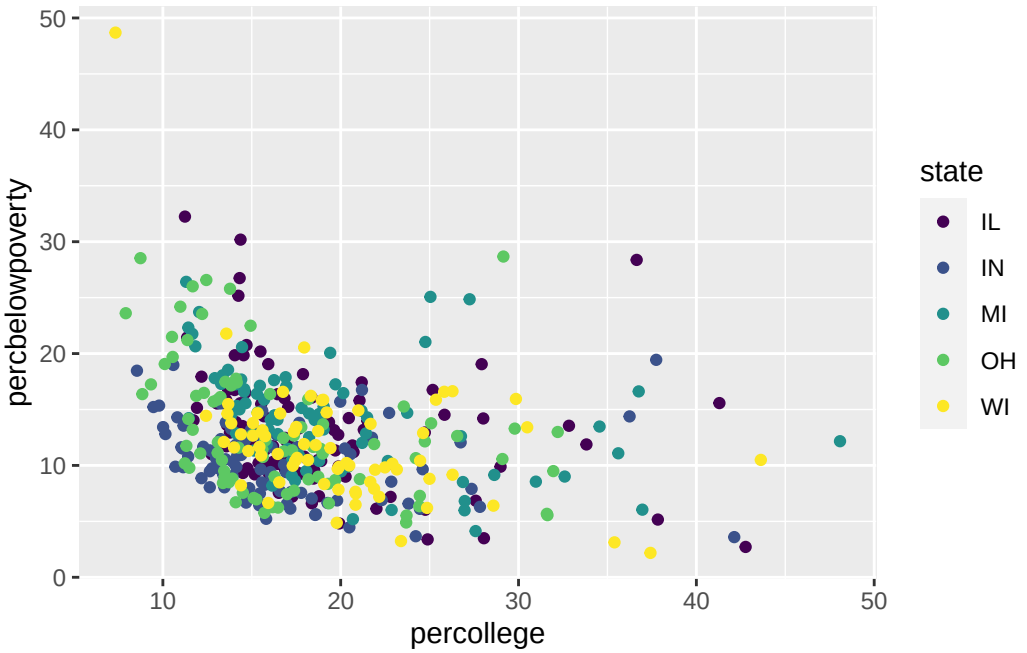
```
ggplot(data = midwest, mapping = aes(x = popdensity)) +
  geom_histogram(binwidth = 10000)
```



The data is greatly skewed to the right. There appears to be outliers between 50000 and 75000 as well as after 75000.

Exercise 2

```
ggplot(data = midwest, mapping = aes(x = percollege, y = percbelowpoverty, color = state))
     geom_point() +
        scale_color_viridis_d()
```



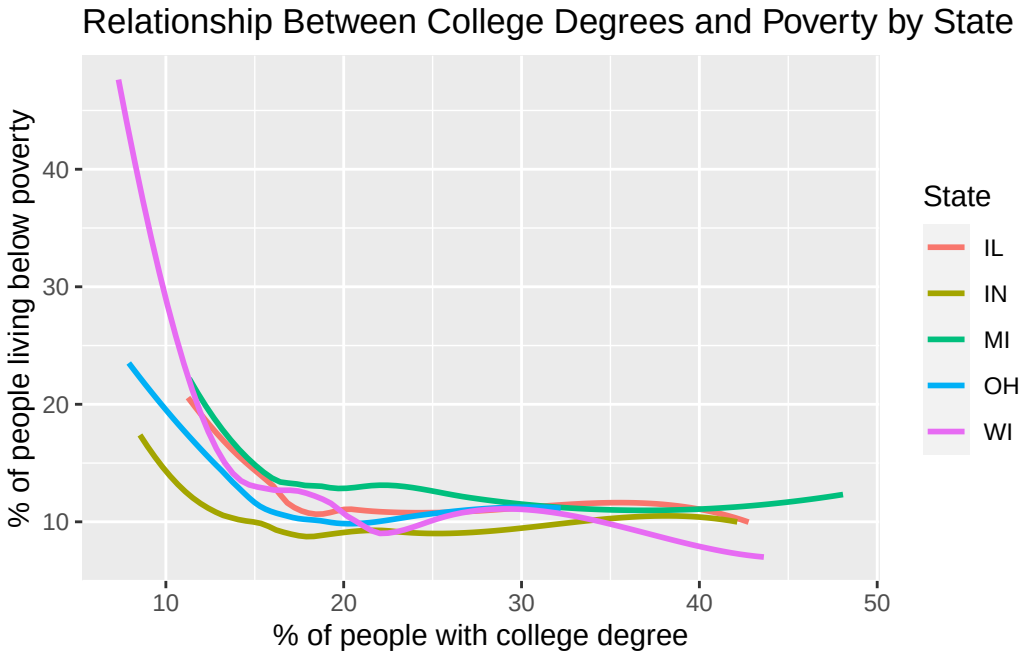
Exercise 3

There is a large clump of states in this graph making it hard to interpret. It seems that Ohio has a large amount of people living below the poverty line and not a lot of people with college degrees. For Illinois, it seems like more people have college degrees due to the purple dots off to the right side of the graph.

Exercise 4

```
ggplot(data = midwest, mapping = aes(x = percollege, y = percbelowpoverty, color = state))
     geom_smooth(se = FALSE) +
  labs(x = "% of people with college degree",
       y = "% of people living below poverty",
       title = "Relationship Between College Degrees and Poverty by State",
       color = "State")
```

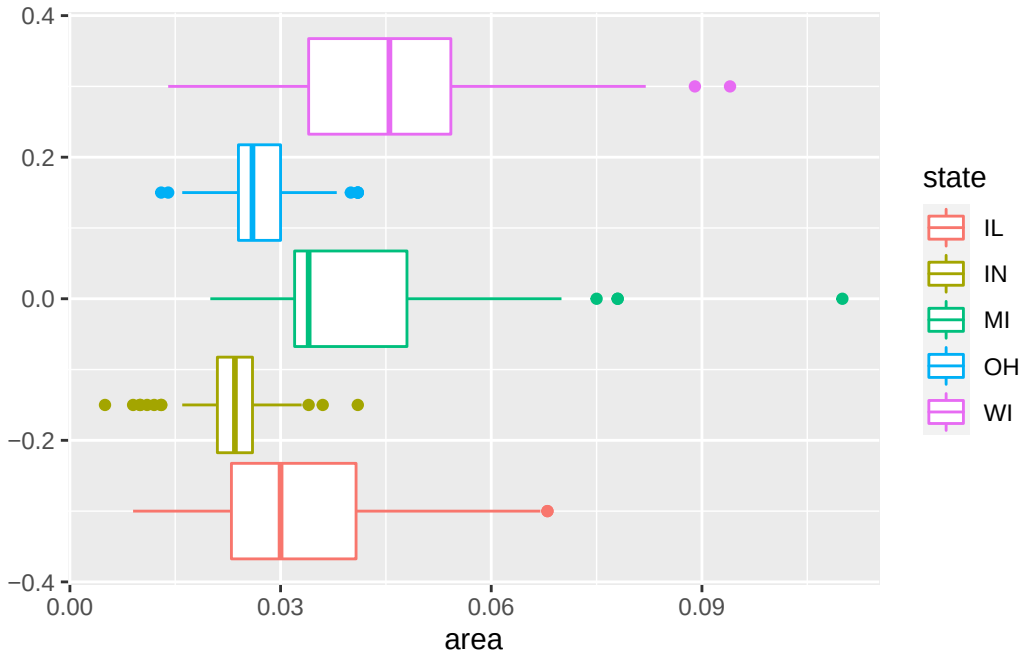
```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



I prefer this plot from exercise 4 because it is easier to read. It is easier to see the trends here by dragging a finger and mapping what is on the X axis to what is on the Y axis.

Exercise 5

```
ggplot(data = midwest, mapping = aes(x = area,color = state)) +
  geom_boxplot()
```

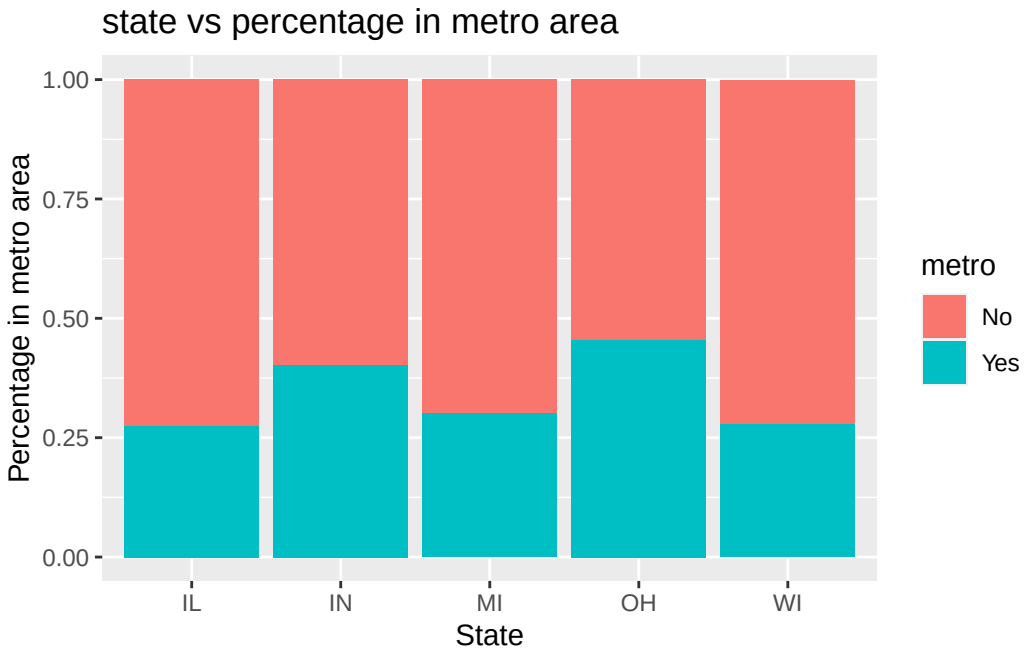


Wisconsin's median is larger than the other states. The red and purple one have the most variability, they are the most spread out.

- Which state has the single largest county? How do you know based on the plot?

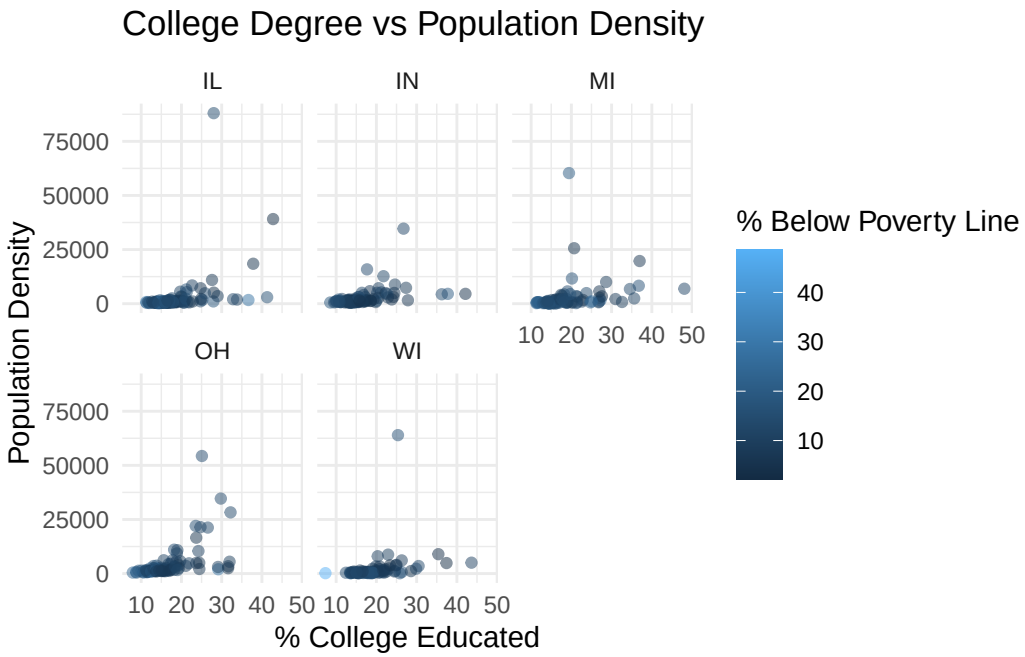
Exercise 6

```
midwest <- midwest |>
  mutate(metro = if_else(inmetro == 1, "Yes", "No"))
     ggplot(midwest,
       aes(x = state, fill = metro)) +
           geom_bar(position = "fill") +
           labs(title = "state vs percentage in metro area", x = "State", y = "Percentage
```



Exercise 7

```
ggplot(midwest, aes(x= percollege, y = popdensity, color= percbelowpoverty)) +
  geom_point(alpha = .5) +
  facet_wrap(~state) +
  labs(title = "College Degree vs Population Density", x = "% College Educated" , y = "Pop
  theme_minimal()
```


College Degree vs Population Density
IL
IN
MI
OH
WI
Population Density
75000
50000
25000
0
10 20 30 40 50
% College Educated
% Below Poverty Line
40
30
20
10