Lab 1 - Data visualization

Karina Marinovich

Load Packages

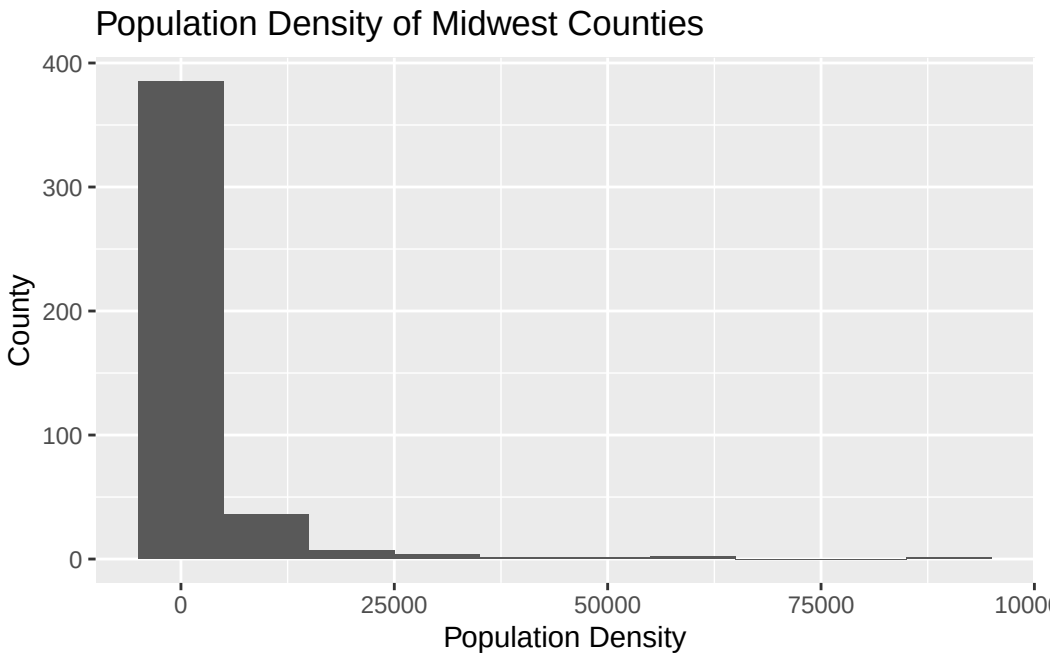
```
library(tidyverse)
```

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

```
library(viridis)
```

Exercise 1

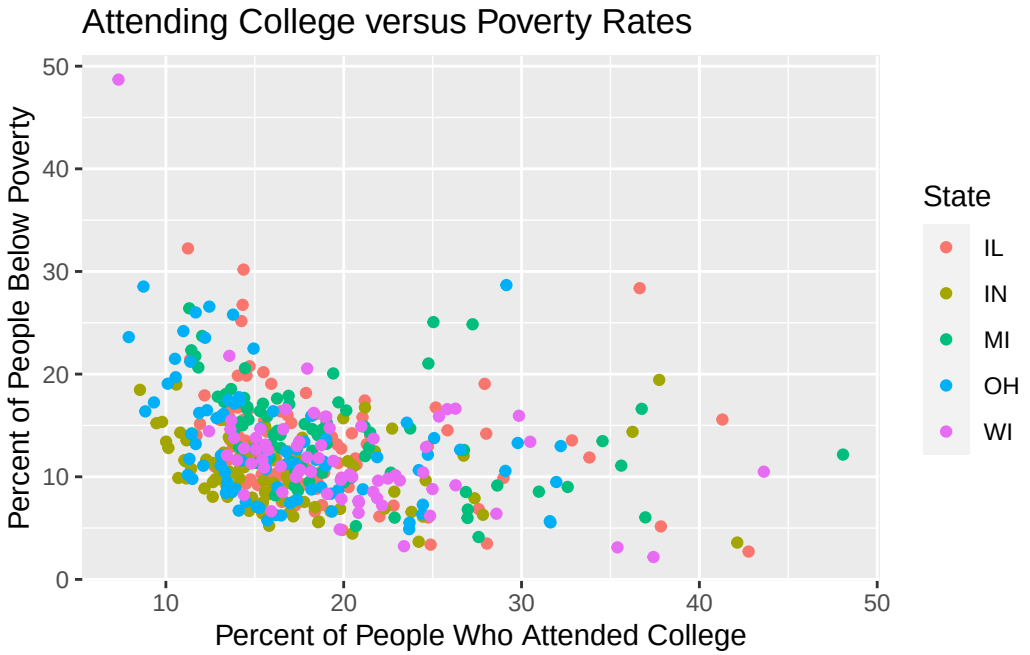
```
ggplot(midwest,
       aes(x = popdensity)) +
  geom_histogram(binwidth = 10000) +
  labs(x = "Population Density", y = "County",
       title = "Population Density of Midwest Counties")
```



The shape of the distribution of Population Density of Midwest Counties is asymmetrical, skewed right. There appears to be an outlier when the population density is about 90,000.

Exercise 2

```
ggplot(midwest, aes(x = percollege, y = percbelowpoverty, color = state)) +
  geom_point() + labs(x= "Percent of People Who Attended College",
                      y = "Percent of People Below Poverty", color = "State",
title = "Attending College versus Poverty Rates") + scale_fill_viridis()
```



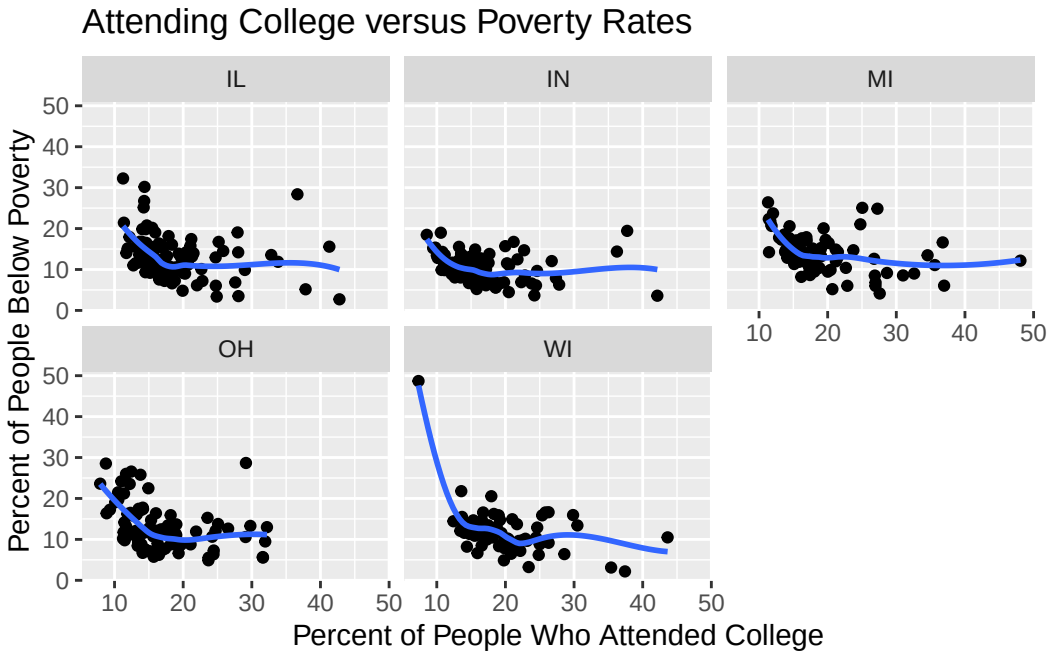
Exercise 3

There is a negative correlation between the percent of people who attended college and percent of people below the poverty line. This suggests that attending college increases one's likelihood of not living in poverty. Illinois shows the weakest correlation, as its data points are considerably scattered. Indiana's data points appear to be the greatest in the percentage of people below poverty. Michigan and Illinois show the highest percentage of people living below poverty, relative to the other states. Michigan holds the greatest outlier for percent of people who attended college at around 48% and Wisconsin has the greatest outlier for percent of people beloow poverty at around 49%. Overall, the trend for the relationship between college attendence and poverty is very similar for these states.

Exercise 4

```
ggplot(midwest, aes(x = percollege, y = percbelowpoverty)) + geom_point() +
  labs(x= "Percent of People Who Attended College",
       y = "Percent of People Below Poverty", color = "State",
       title = "Attending College versus Poverty Rates") +
  facet_wrap(~state) + geom_smooth(se = FALSE)
```

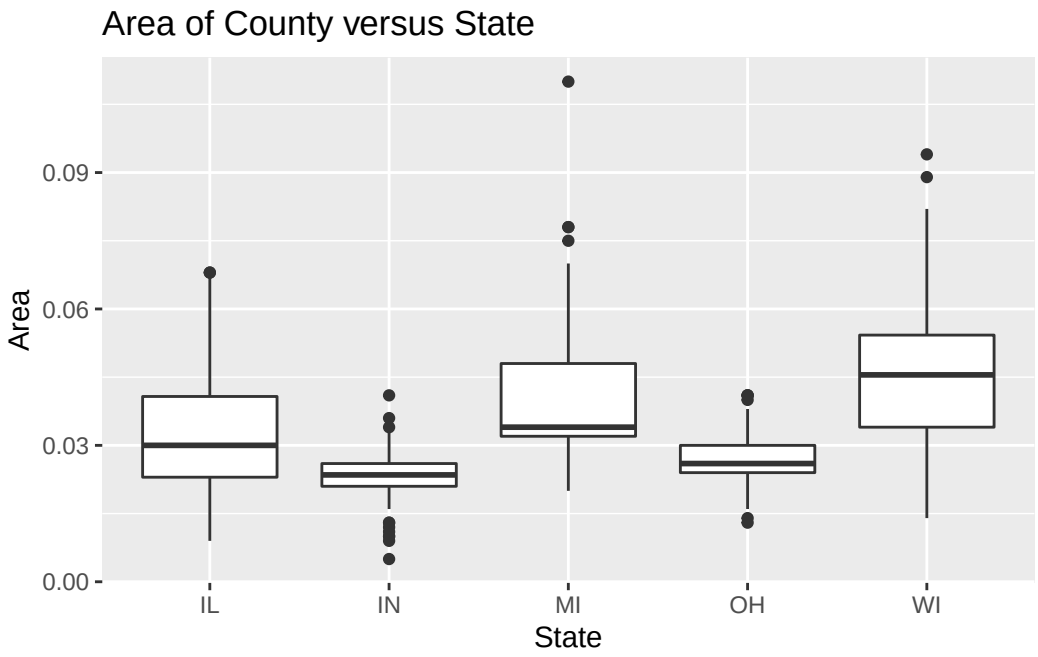
```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



I prefer the plot in excercise 4 because the overlapping, cluttered, different colored data points in ex 2 made it difficult to compare the trends of different states. This plot separates the states' data points, making it easier to compare trends side by side. I also appreciate the lines, which really help me visualize the data trends.

Exercise 5

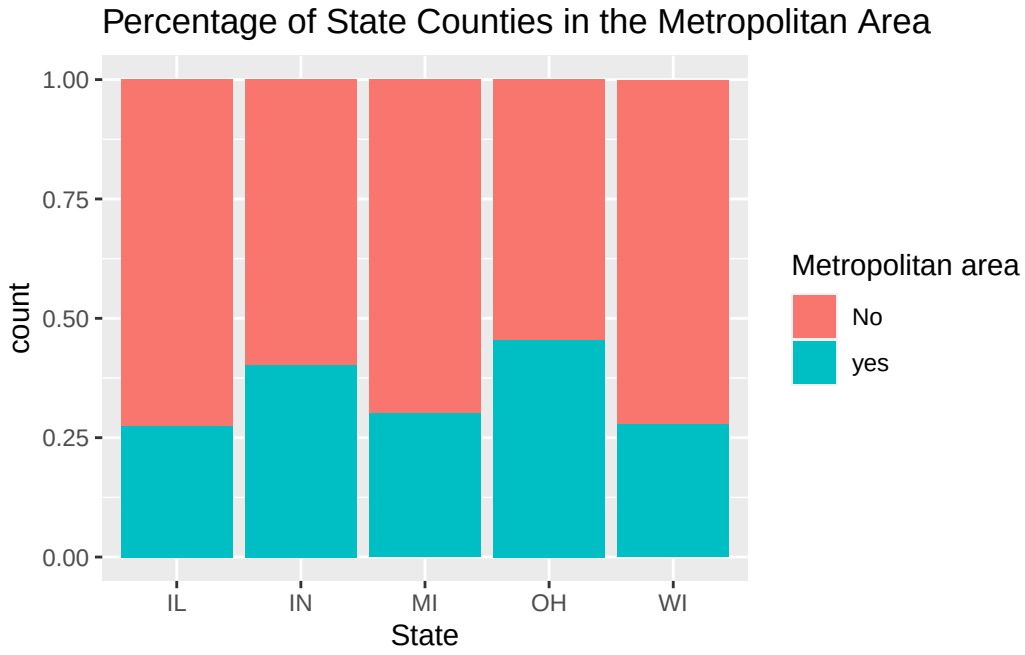
```
ggplot(midwest, aes(x = state, y= area)) + geom_boxplot() +
  labs(x = "State", y = "Area", title = "Area of County versus State")
```



Some states do tend to have geographically larger counties. Wisconsins' counties appear to be the largest because its median area is the greatest relative to the other states. Indiana's median is the lowest, indicating that its counties are typically the smallest in area. Michigan is home to the single largest county because it contains an outlier point higher than any other plot. Furthermore, Indiana holds the smallest outlier county. The distributions are fairly symmetric, except for Michigan's, which is skewed right.

Exercise 6

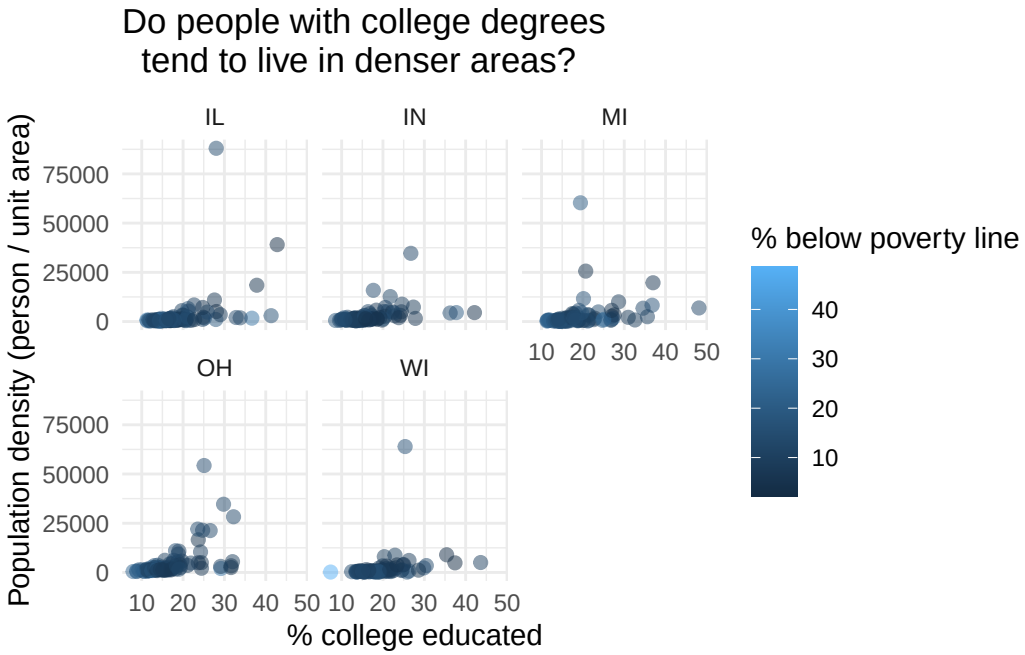
```
midwest <- midwest |>
  mutate(metro = if_else(inmetro ==1, "yes", "No"))
ggplot(midwest, aes(x = state, fill = metro)) +
  geom_bar(position = "fill") +
  labs(x = "State", fill = "Metropolitan area",
       title = "Percentage of State Counties in the Metropolitan Area")
```



For all of these Midwest states, there is a lower percentage of state counties metropolitan areas than not. Illinois, Michigan, and Wisconsin all have the lowest percentages of counties in metropolitan areas, while Ohio has the greatest and Indiana follows behind.

Exercise 7

```
ggplot(midwest,
       aes(x = percollege, y = popdensity, color = percbelowpoverty)) +
  geom_point(size = 2, alpha = 0.5) +
  facet_wrap(~state) +
  labs(x = "% college educated", y = "Population density (person / unit area)",
  color = "% below poverty line", title = "Do people with college degrees
  tend to live in denser areas?") + scale_fill_viridis() + theme_minimal()
```


Do people with college degrees
tend to live in denser areas?
IL
IN
MI
OH
WI
Population density (person / unit area)
75000
50000
25000
0
10 20 30 40 50
% college educated
% below poverty line
40
30
20
10