Lab 1 - Data visualization

Carli Olsen

Load Packages

```
library(tidyverse)
```

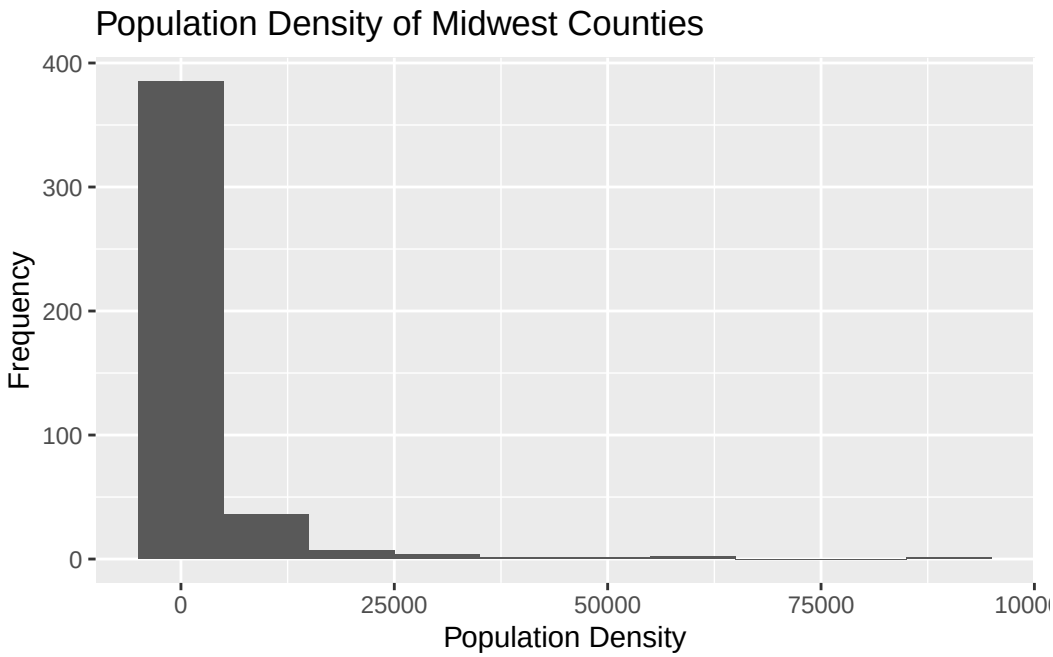
```
Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
had status 1
```

```
library(viridis)
```

Exercise 1

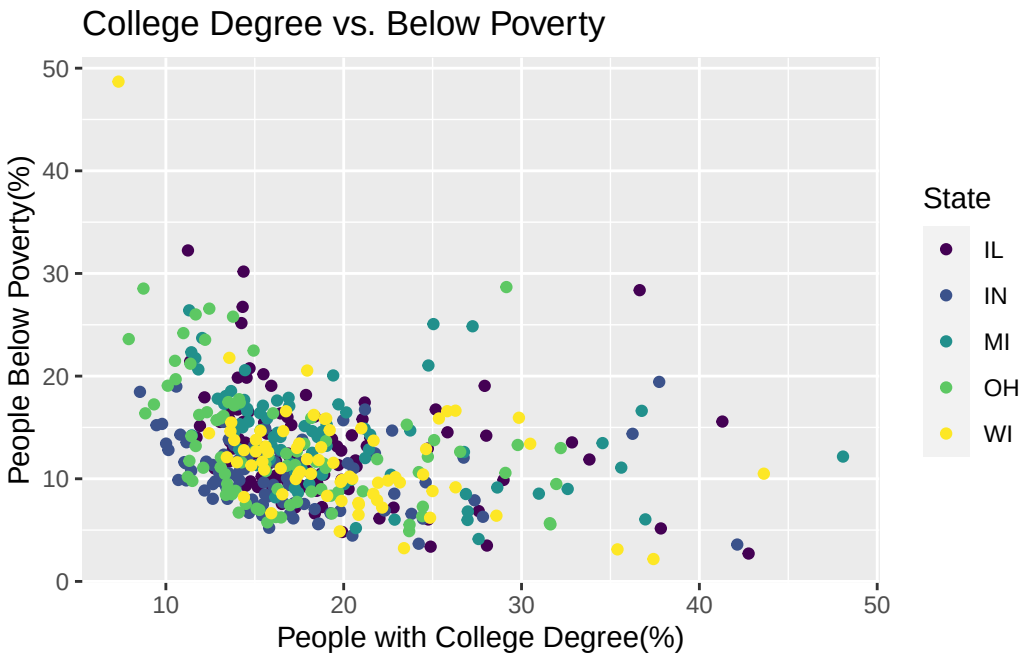
This histogram has a right skew. There appears to be an outlier around 90,000 because it is very far away from the median and there is a low frequency.

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



Exercise 2

```
ggplot(midwest, aes(x=percollege, y = percbelowpoverty)) +
  geom_point(aes(color = state)) +
  labs(title = "College Degree vs. Below Poverty",x = "People with College Degree(%)",y =
  scale_color_viridis_d()
```



Exercise 3

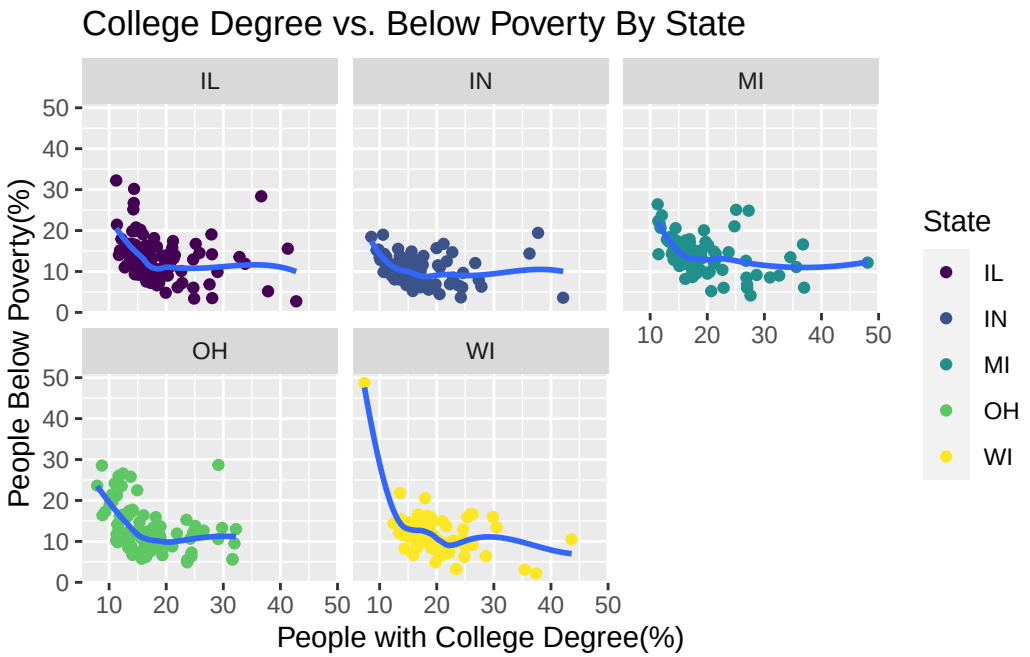
It appears that in most of the states,if you have a higher percentage of people with a college degree, there is a lower percentage of people below poverty. There is one statistic from Wisconsin that has a bit of an outlier with a very high percentage of people below poverty. In Ohio it seems that there is less people with college degrees. Illinois tends to have a high percentage of people in poverty regardless of the percentage of people with college degress.

Exercise 4

I prefer this plot because it is much easier to read because there is less overlap amongst dots.

```
ggplot(midwest, aes(x=percollege, y = percbelowpoverty)) +
  geom_point(aes(color = state)) +
  labs(title = "College Degree vs. Below Poverty By State",x = "People with College Degree
  scale_color_viridis_d() +
  facet_wrap(~state) +
  geom_smooth(se = FALSE)
```

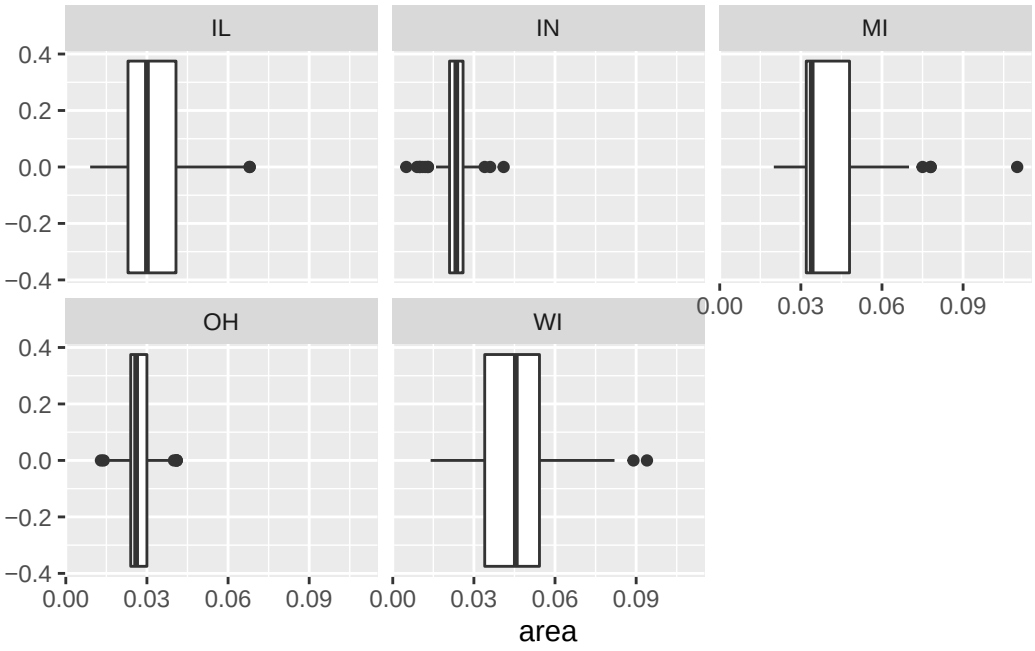
```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Exercise 5

Ohio and Illinois have the most consistent size of counties since their boxes and tails are the smallest. Michigan has the largest county because it contains the dot furthest to the right and it is “1

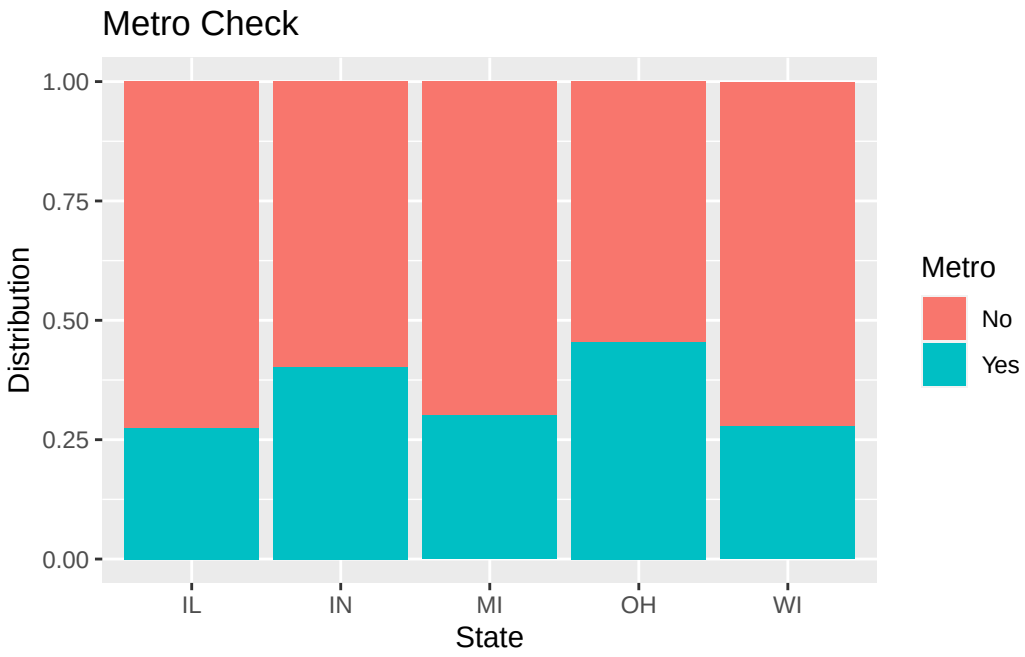
```
ggplot(midwest, aes(x = area)) +
geom_boxplot() +
  facet_wrap(~state)
```



Exercise 6

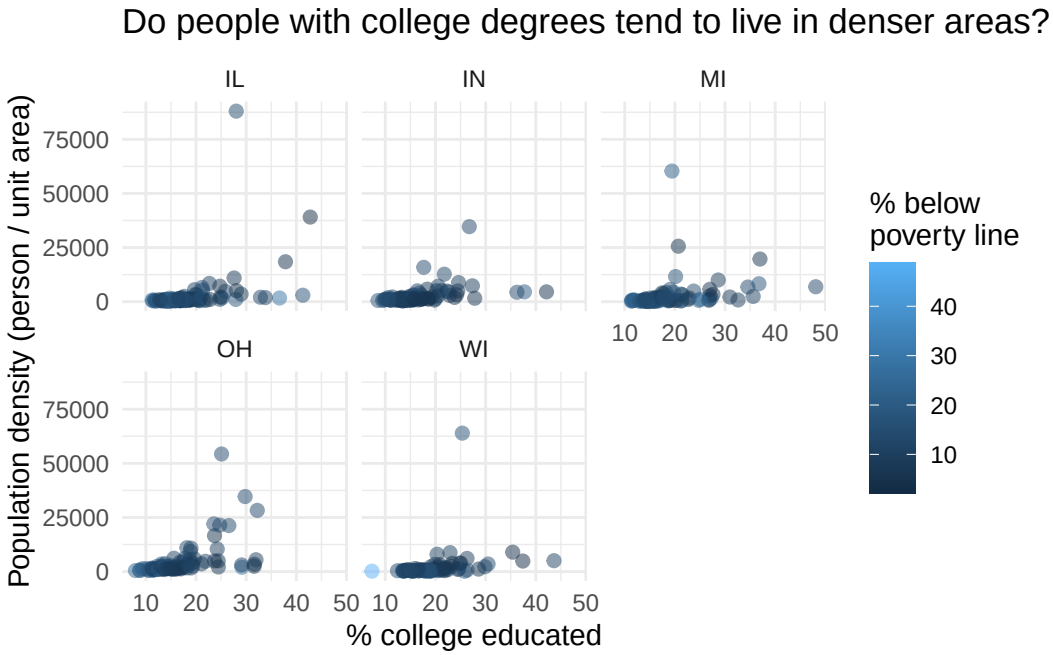
Indiana and Ohio have the highest percentage of counties located in a metropolitan area.

```
midwest <- midwest |>
  mutate(metro = if_else(inmetro == 1, "Yes", "No"))
ggplot(midwest, aes(x = state)) +
  geom_bar(aes(fill = metro), position = "fill") +
  labs(title = "Metro Check", x = "State", y = "Distribution", fill = "Metro")
```



Exercise 7

```
ggplot(midwest, aes(x = percollege, y = popdensity)) +
  geom_point(alpha = 0.5, size = 2, aes(color = percbelowpoverty)) +
  labs(title = "Do people with college degrees tend to live in denser areas?",x = "% colle
  facet_wrap(~state) +
  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