Lab 1 - Data visualization

Nicolas Pardo

Load Packages

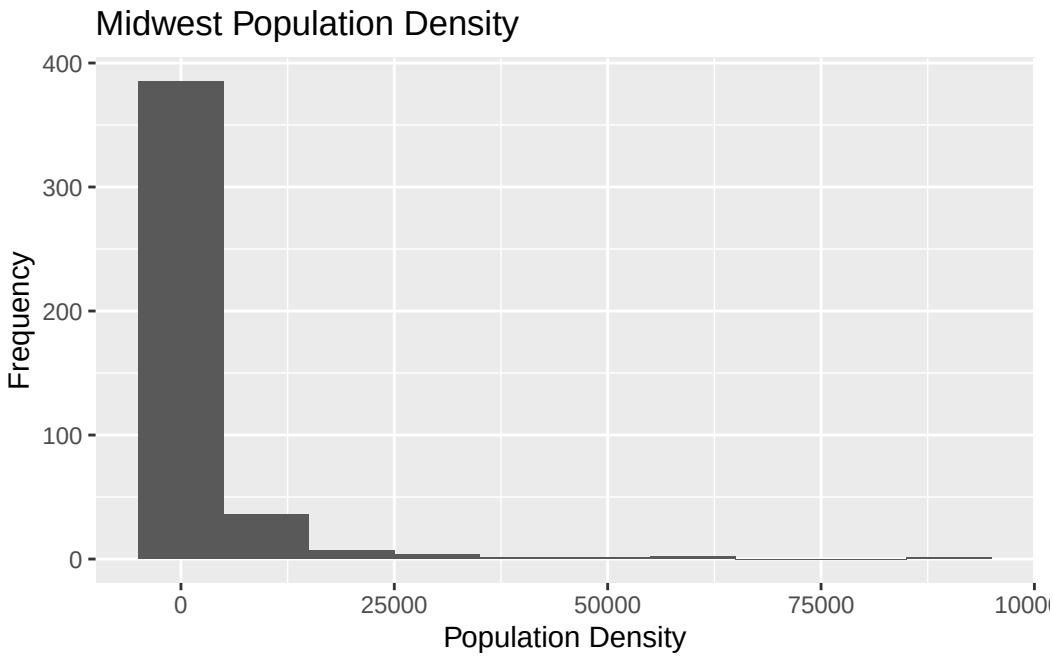
```
library(tidyverse)
```

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

```
library(viridis)
```

Exercise 1

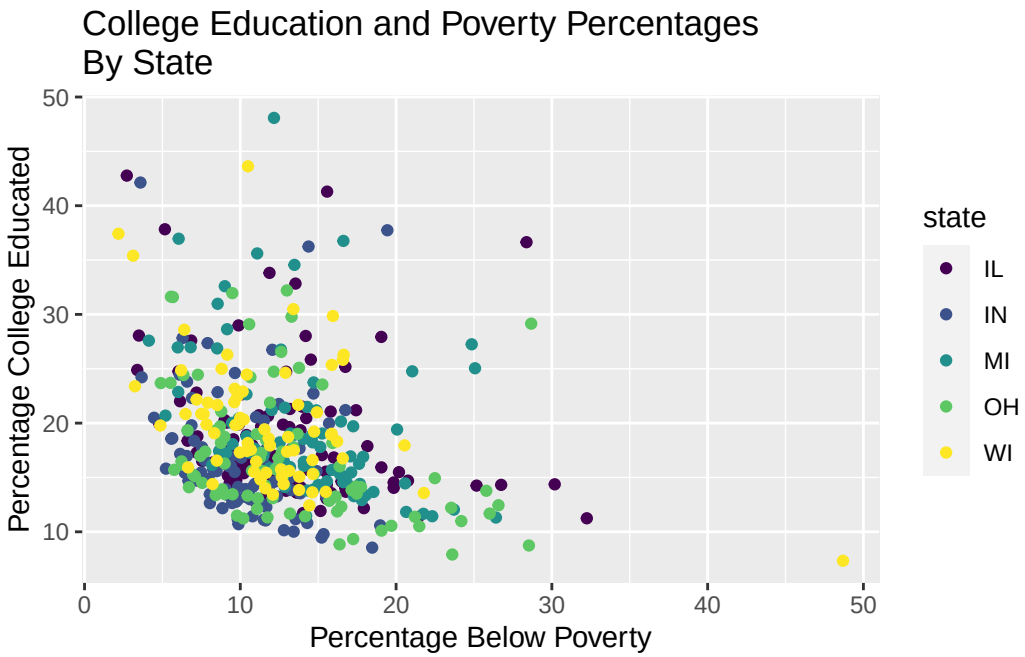
```
#|label: ex-1
ggplot(midwest, aes(x = popdensity))+
  geom_histogram(binwidth = 10000)+
  labs(title = "Midwest Population Density", y = "Frequency", x = "Population Density")
```



The distribution is unimodal and right skewed. There is an outlier around 90000.

Exercise 2

```
#|label: ex-2
ggplot(midwest, aes(x = percbelowpoverty, y = percollege, color = state))+
  geom_point()+
  labs(title = "College Education and Poverty Percentages \nBy State", x = "Percentage Bel
scale_color_viridis_d()
```



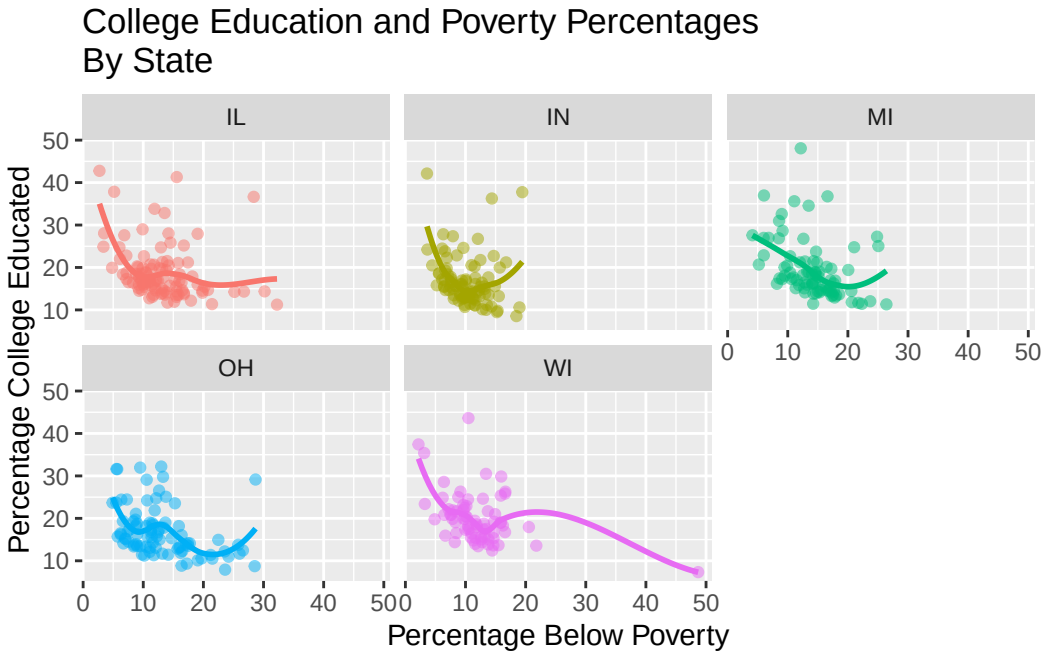
Exercise 3

There is a large concentration of points around 10% below poverty and 15% college educated for all of the states. WI has one county with the highest poverty percentage of all the states. MI has one county with the highest percentage of college educated adults. Overall the states have very similar distributions of poverty percentages and college educated percentages.

Exercise 4

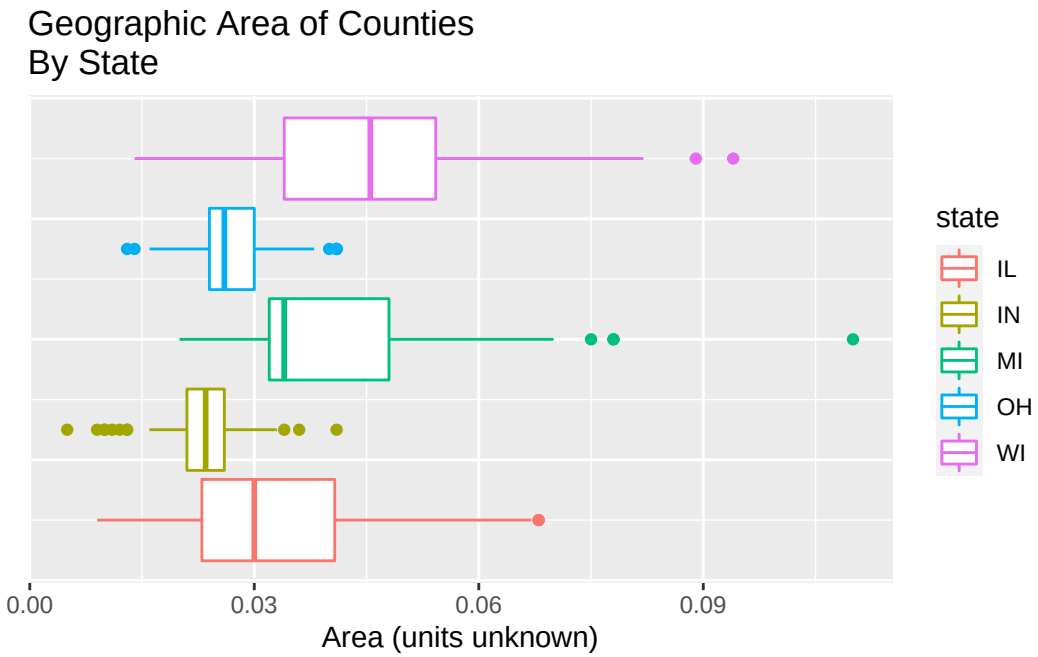
```
#|label: ex-4
ggplot(midwest, aes(x = percbelowpoverty, y = percollege, color = state))+
  geom_smooth(se = FALSE, show.legend = F)+
  geom_point(show.legend = F, alpha = .5)+
  facet_wrap(~state)+
  labs(title = "College Education and Poverty Percentages \nBy State", x = "Percentage Bel
```

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



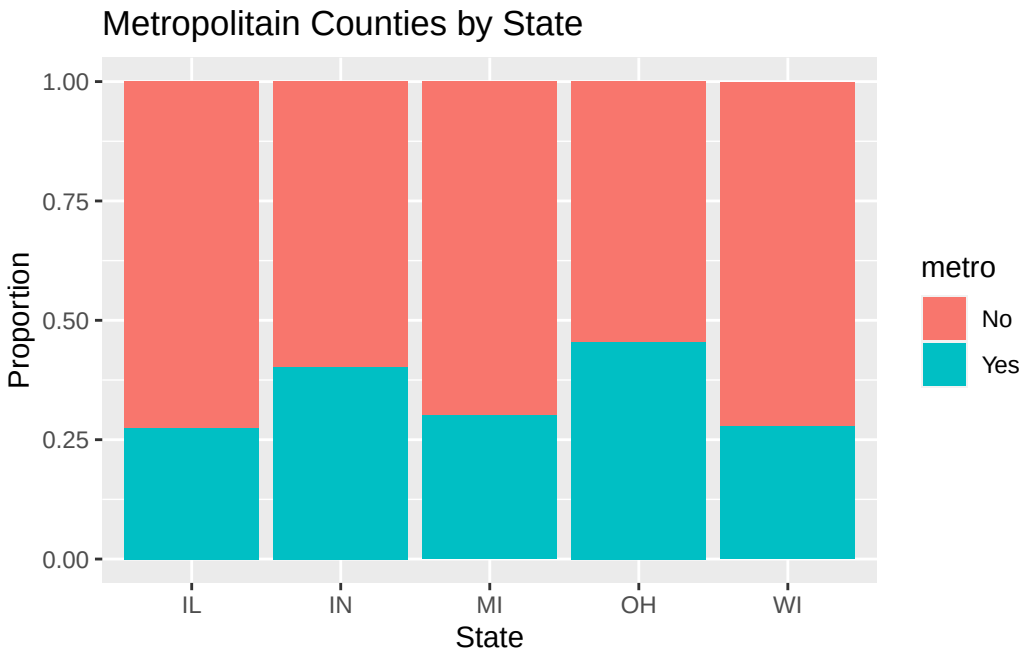
I prefer this visualization because the states are not overlapping and you can see more of the distribution as well as geom_smooth giving more information. ## Exercise 5

```
#|label: ex-5
ggplot(midwest, aes(x = area, color = state))+
  geom_boxplot()+
  labs(title = "Geographic Area of Counties \nBy State", x = "Area (units unknown)")+
  theme(axis.ticks.y = element_blank(), axis.text.y = element_blank())
```



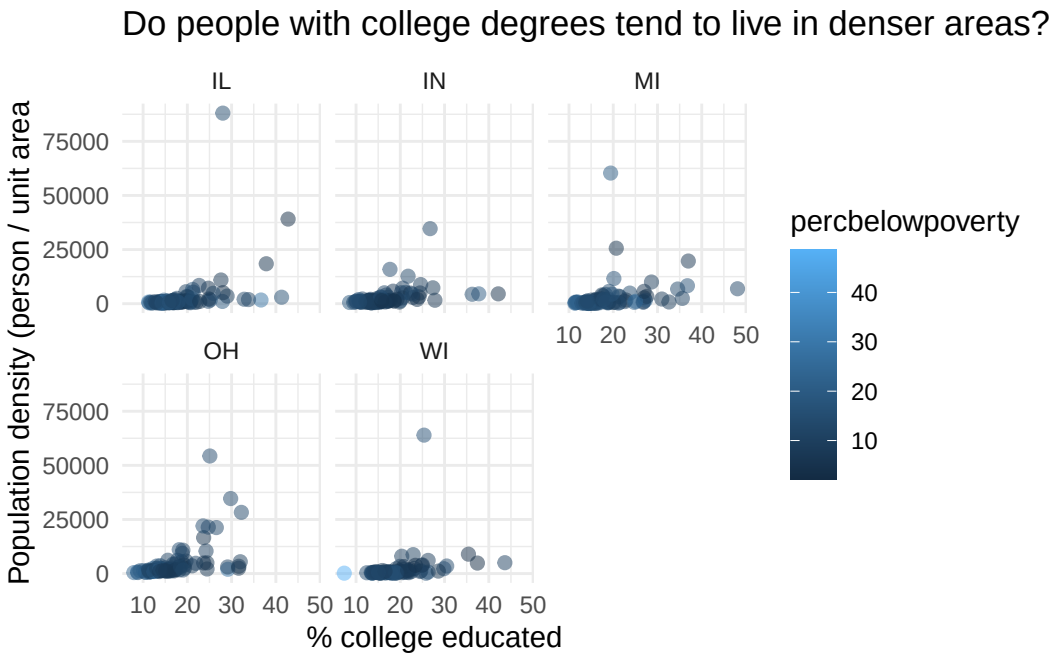
The single largest county is in MI as seen by the green dot on the far right. This plot shows the distributions, medians, and outliers of each states area. WI has the largest median, while the other states are around 0.03. IN and OH have the smallest ranges. ## Exercise 6

```
#|label: ex-6
midwest <- midwest |>
  mutate(metro = if_else(inmetro == 1, "Yes", "No"))
ggplot(midwest, aes(x = state, fill = metro))+
  geom_bar(position = "fill")+
  labs(title = "Metropolitain Counties by State", x = "State", y = "Proportion")
```



This chart shows the proportion of counties in metro areas. OH has the most counties in metro areas while IL and WI have similar, low, levels of counties in metro areas. MI has slights more metro counties than IL and WI. ## Exercise 7

```
#|label: ex7
ggplot(midwest, aes(x = percollege, y = popdensity, color = percbelowpoverty))+
  geom_point(size = 2, alpha = 0.5)+
  facet_wrap(~state)+
  labs(title = "Do people with college degrees tend to live in denser areas?", y = "Popula
  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
percbelowpoverty
40
30
20
10