

# Lab 1 - Data visualization

Lauren Valle

## Load Packages

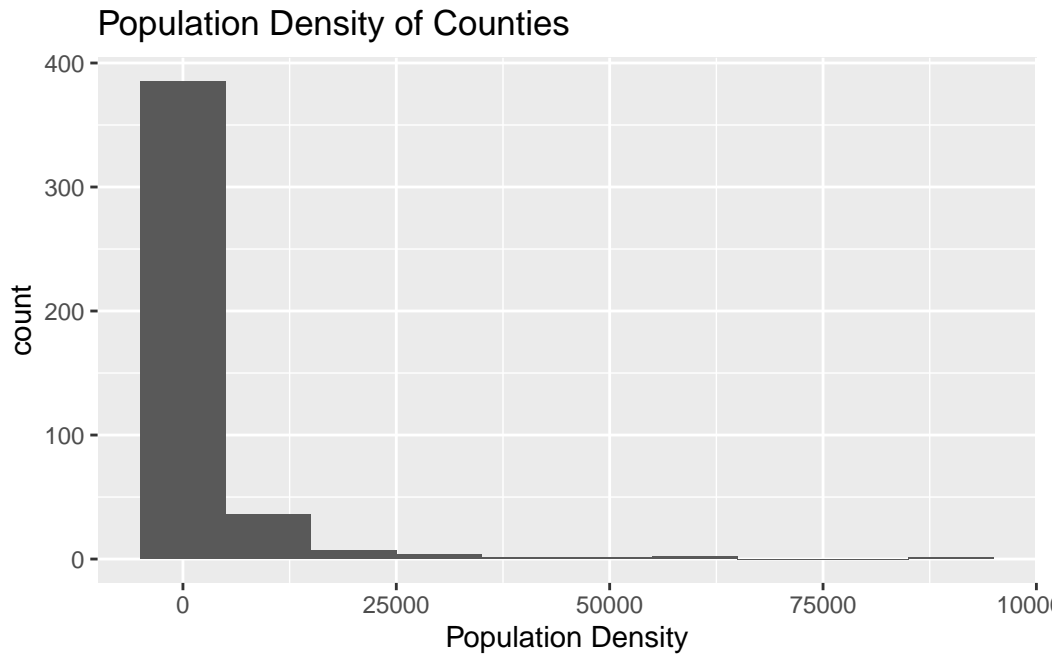
```
library(tidyverse)
```

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

```
library(viridis)
```

## Exercise 1

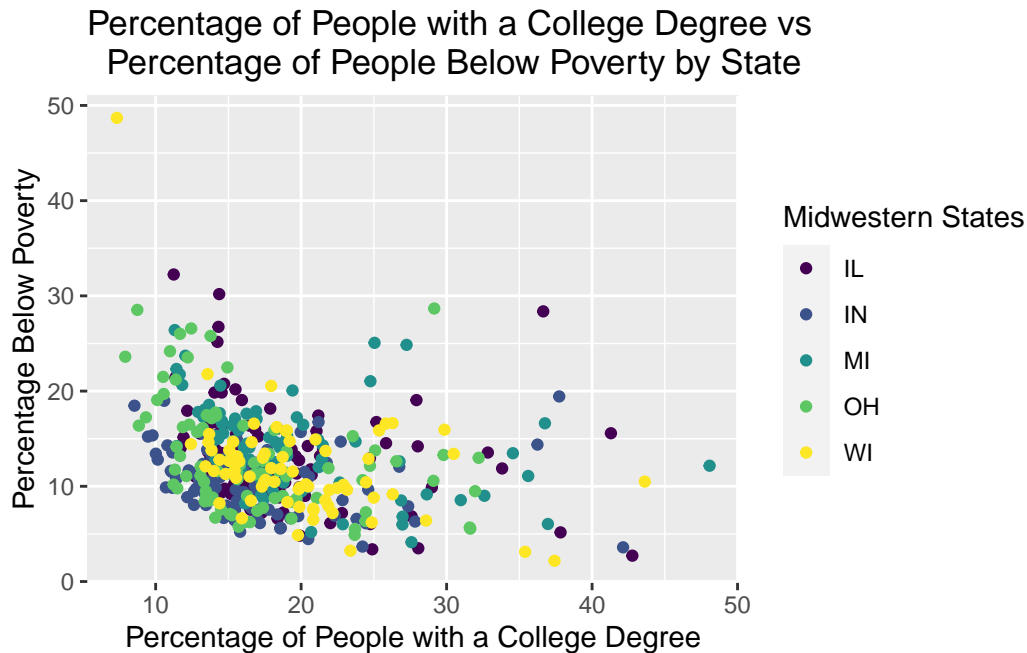
```
ggplot(midwest)+  
  geom_histogram( aes(popdensity), binwidth = 10000)+  
  labs(  
    x = "Population Density",  
    y = "count",  
    title = "Population Density of Counties"  
  )
```



This histogram is right skewed. It looks like there is an outlier in between  $x=7500$  and  $x = 100000$

## Exercise 2

```
ggplot(data = midwest,
       mapping = aes(x = percollege, y = percbelowpoverty, color = state)) +
  geom_point() +
  labs(
    x = "Percentage of People with a College Degree",
    y = "Percentage Below Poverty",
    title = "Percentage of People with a College Degree vs
Percentage of People Below Poverty by State",
    color = "Midwestern States"
  ) +
  scale_color_viridis_d()
```



### Exercise 3

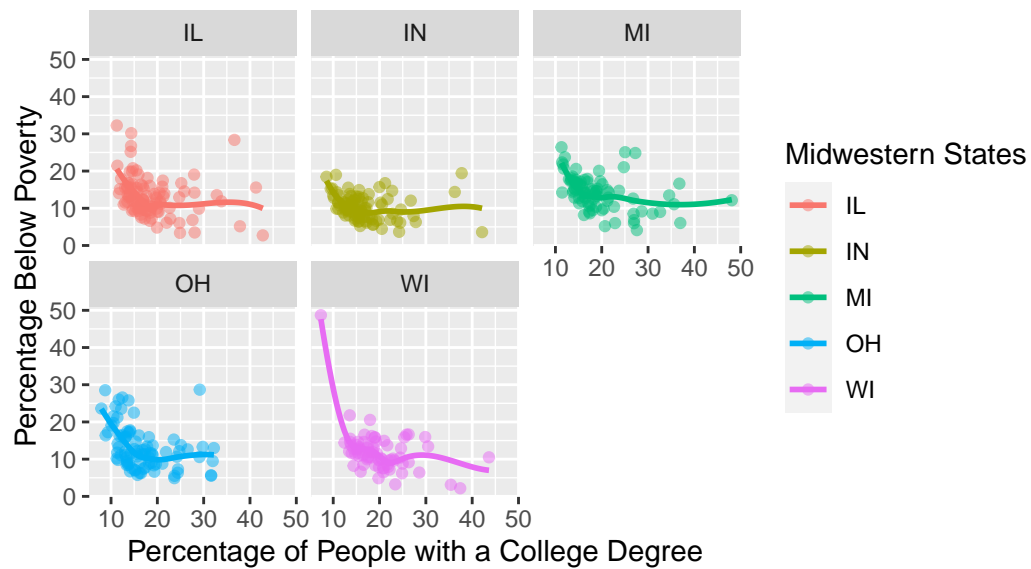
This is a general trend among the states included in the dataset that those below the poverty line are less likely to obtain a college degree.

### Exercise 4

```
ggplot(data = midwest,
       mapping = aes(x = percollege, y = percbelowpoverty, color = state)) +
  geom_smooth(se = FALSE)+
  geom_point(alpha = 0.5)+
  labs(
    x = "Percentage of People with a College Degree",
    y = "Percentage Below Poverty",
    title = "Percentage of People with a College Degree vs
Percentage of People Below Poverty by State",
    color = "Midwestern States"
  )+
  facet_wrap("state")
```

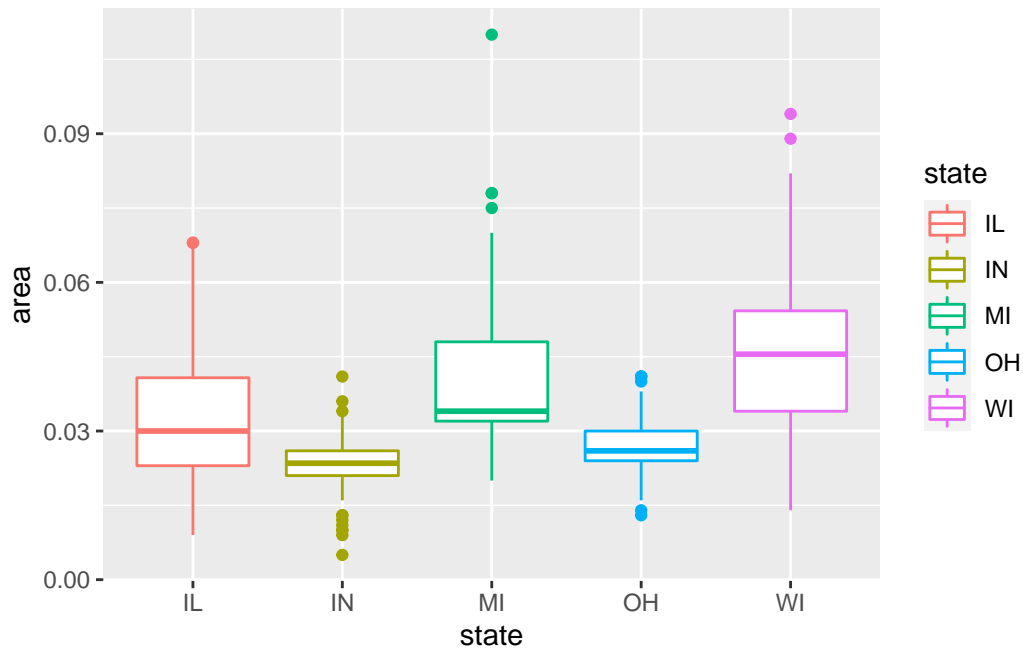
`geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Percentage of People with a College Degree vs Percentage of People Below Poverty by State



### Exercise 5

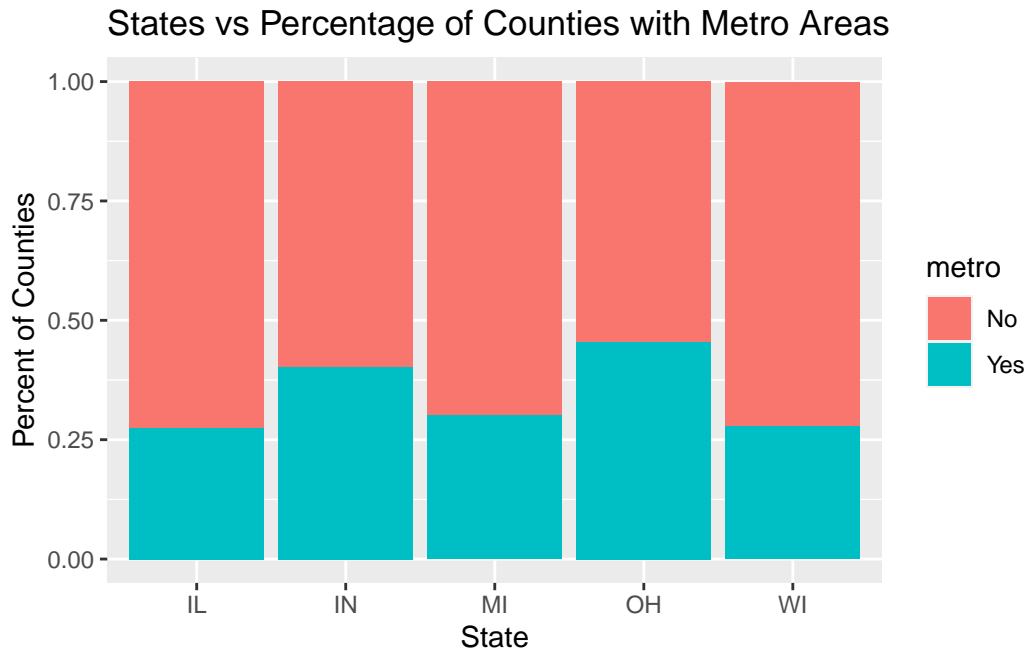
```
ggplot(data = midwest,  
       mapping = aes(x = state, y = area, color = state)) +  
  geom_boxplot()
```



From this side by side box plot we can see that Michigan has the largest single county (as show by the single green dot much higher than the rest of the data set). I also notice that Michigan and Wisconsin tend to have counties larger in area while Indiana and Ohio tend to have counties that are on the smaller side (in comparison).

## Exercise 6

```
midwest <- midwest |>
  mutate(metro = if_else(inmetro == 1, "Yes", "No"))
ggplot(data = midwest, mapping = aes(x = state, fill = metro))+
  geom_bar(position = "fill")+
  labs(
    x = "State",
    y = "Percent of Counties",
    title = "States vs Percentage of Counties with Metro Areas",
    color = "Metro"
  )
```



Ohio has the highest percent of counties located in a metropolitan area (in this dataset) and Wisconsin and Illinois are tied for the least percent (in this dataset)

## Exercise 7

```
ggplot(data = midwest,
       mapping = aes(x = percollege, y = popdensity, color = percbelowpoverty)) +
  geom_point(size = 2, alpha = 0.5)+
  facet_wrap(~state)+
  labs(
    x = "Percent College Educated",
    y = "Population Density (person/unit area)",
    title = "Do People with College Degrees Live in Denser Areas?",
    color = "% below \npoverty line")+
  theme_minimal()
```

## Do People with College Degrees Live in Denser Areas?

