

# Lab 1 - Data visualization

Dennis Park

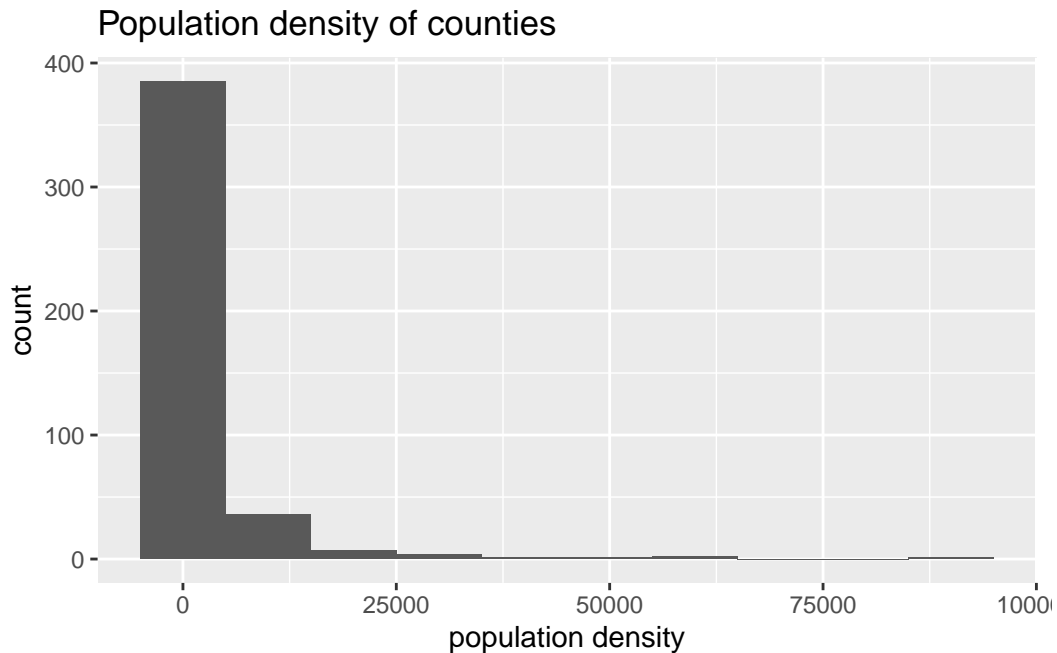
## Load Packages

```
library(tidyverse)
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.)

```
# label:population-density-of-counties
ggplot(data = midwest,
       aes(x = popdensity)) +
  geom_histogram(binwidth = 10000) +
  labs(x = "population density",
       y = "count",
       title = "Population density of counties")
```

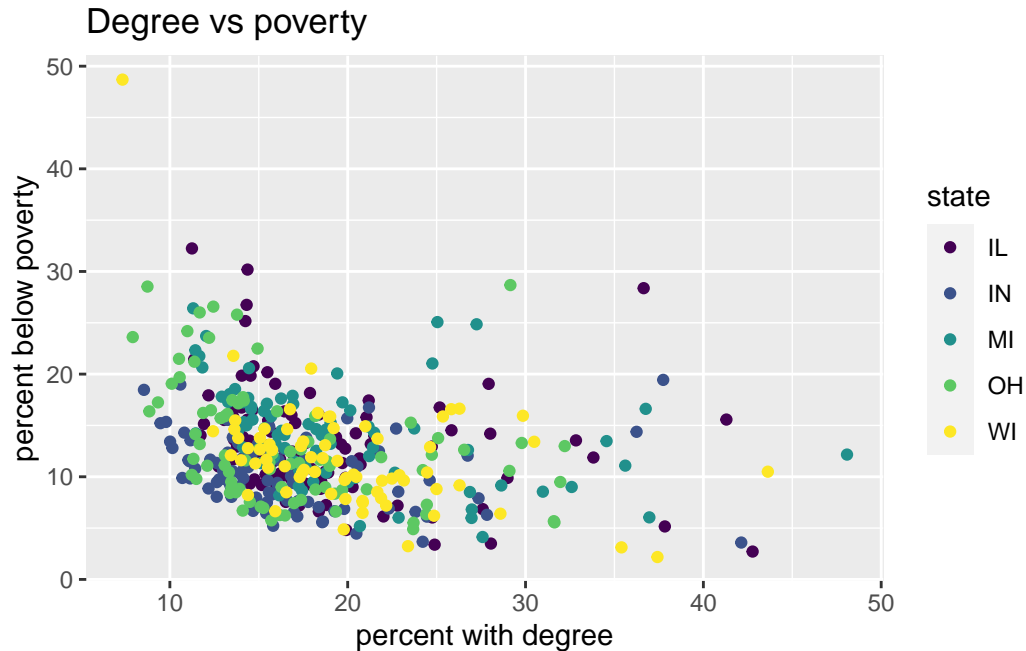


The shape is skewed right, and there are outliers as there extreme data represented on the right side.

## Exercise 2

(Type your answer to Exercise 2 here. Add code chunks as needed. Don't forget to label your code chunk. Do not use spaces in code chunk labels.)

```
# label:percollege-vs-percentbelowpoverty
ggplot(data = midwest,
       aes(x=percollege, y =percbelowpoverty, color = state)) +
  geom_point() +
  labs(x = "percent with degree",
       y = "percent below poverty",
       title = "Degree vs poverty") +
  scale_color_viridis_d()
```



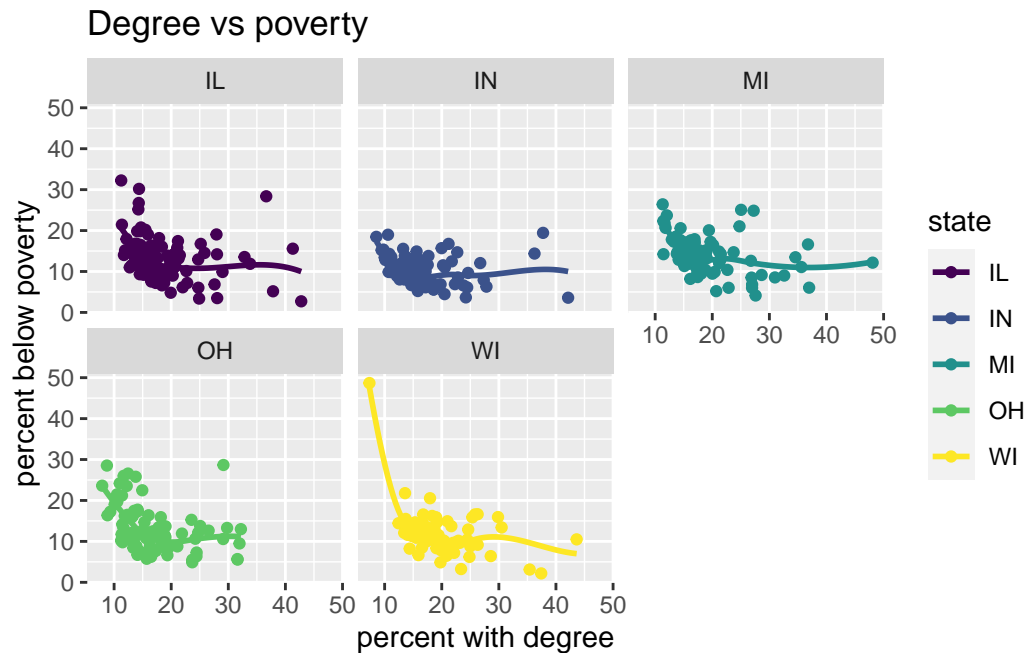
### Exercise 3

(Type your answer to Exercise 3 here. Add code chunks as needed. Don't forget to label your code chunk. Do not use spaces in code chunk labels.)

In most states, there is a great number of data points where both percent below poverty and percent with degree is between 10-30. The states have a similar overall shape of data but states like OH and IL have some extreme points of data.

### Exercise 4

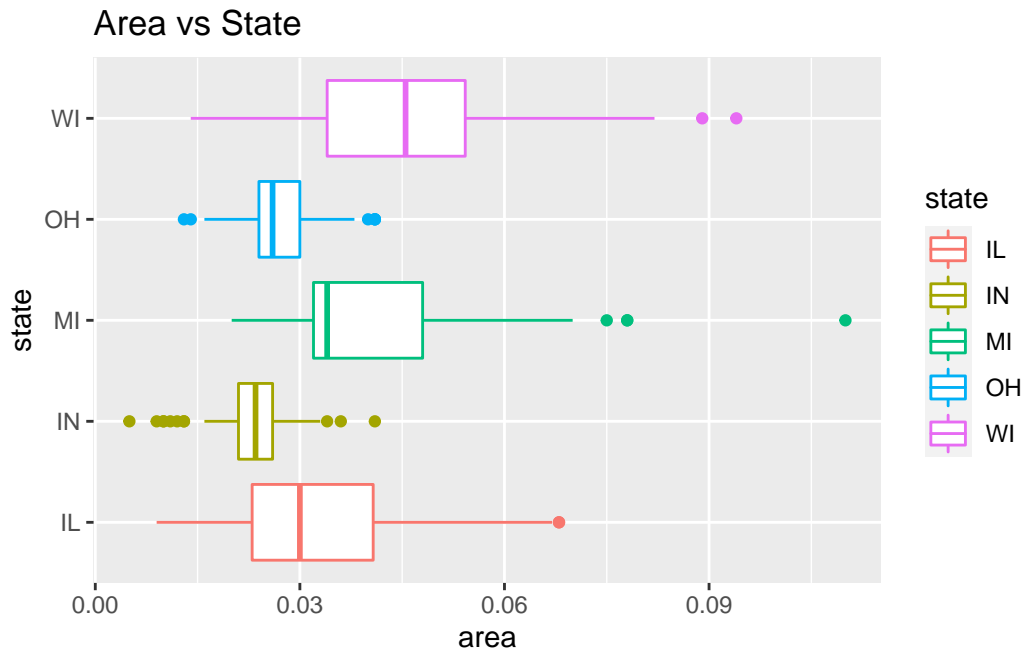
```
# label:Same-as-#3-with-geom-smooth
ggplot(data = midwest,
       aes(x=percollege, y=percbelowpoverty, color = state))+
  geom_point() +
  labs(x = "percent with degree",
       y = "percent below poverty",
       title = "Degree vs poverty")+
  scale_color_viridis_d()+
  facet_wrap("state")+
  geom_smooth(se=FALSE)
```



I prefer this plot instead of Ex 2 because you can focus on each individual state better. You are able to compare them side by side and with the `geom_smooth` function, you can see the direction of the data.

### Exercise 5

```
# label:Area-of-county-based-on-state
ggplot(data = midwest,
       aes(x=area, y =state, color = state))+
  geom_boxplot() +
  labs(x = "area",
       y ="state",
       title = "Area vs State")
```

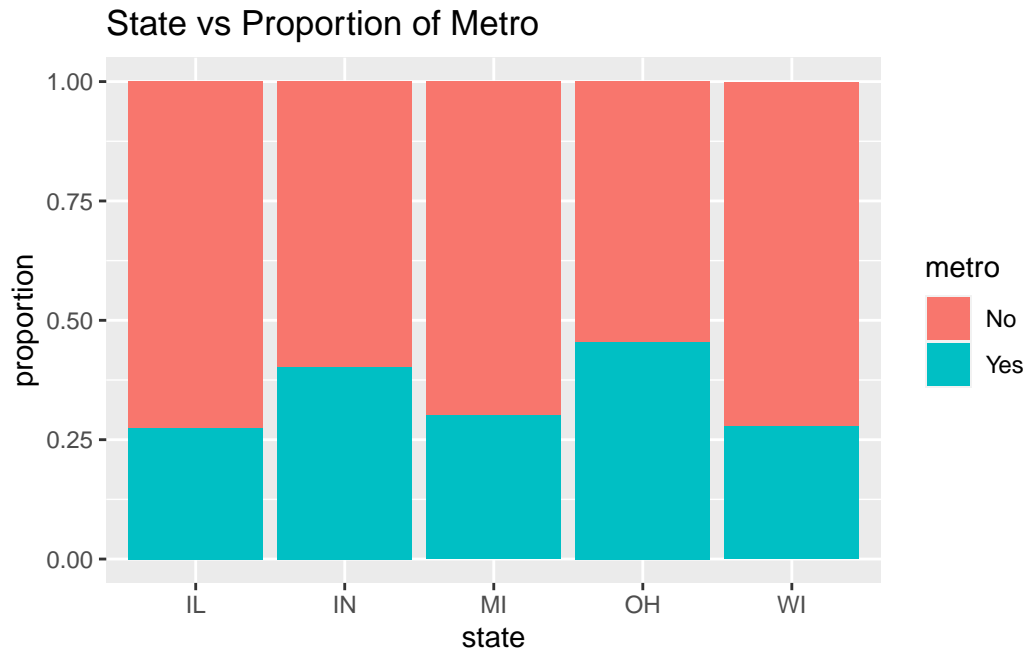


Every state has a median roughly between 0.02 and 0.05. The state WI has the highest median and state IN has the lowest. The state with the single largest county is MI. You can tell by the one extreme data point that MI has that is has a higher area (most right) than any other point. The area of that point is roughly 0.11.

## Exercise 6

```
# label:proportion-of-metro-to-state
midwest <- midwest |>
  mutate(metro = if_else(inmetro == 1, "Yes", "No"))

ggplot(data = midwest,
  aes(x = state, fill = metro)) + geom_bar(position = "fill") +
  labs(x = "state",
    y = "proportion",
    title = "State vs Proportion of Metro")
```



## Exercise 7

```
# label:percent-college-educated-vs-population-density-per-state
ggplot(data = midwest,
       aes(x=percollege, y = popdensity, color = percbelowpoverty))+
  geom_point(alpha = 0.5, size = 2) + theme_minimal() +
  labs(x = "% college educated", y = "population density (person / unit area)") +
  facet_wrap("state")
```

