

# Lab 1 - Data visualization

Abel Abadi

## Load Packages

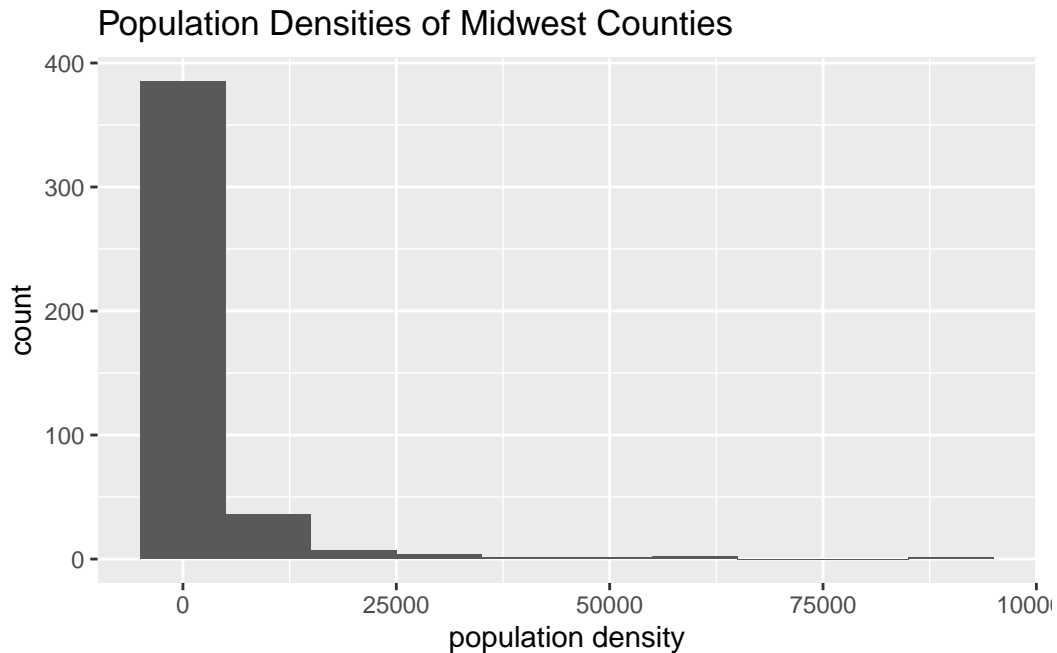
```
library(tidyverse)
```

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

```
library(viridis)
```

## Exercise 1

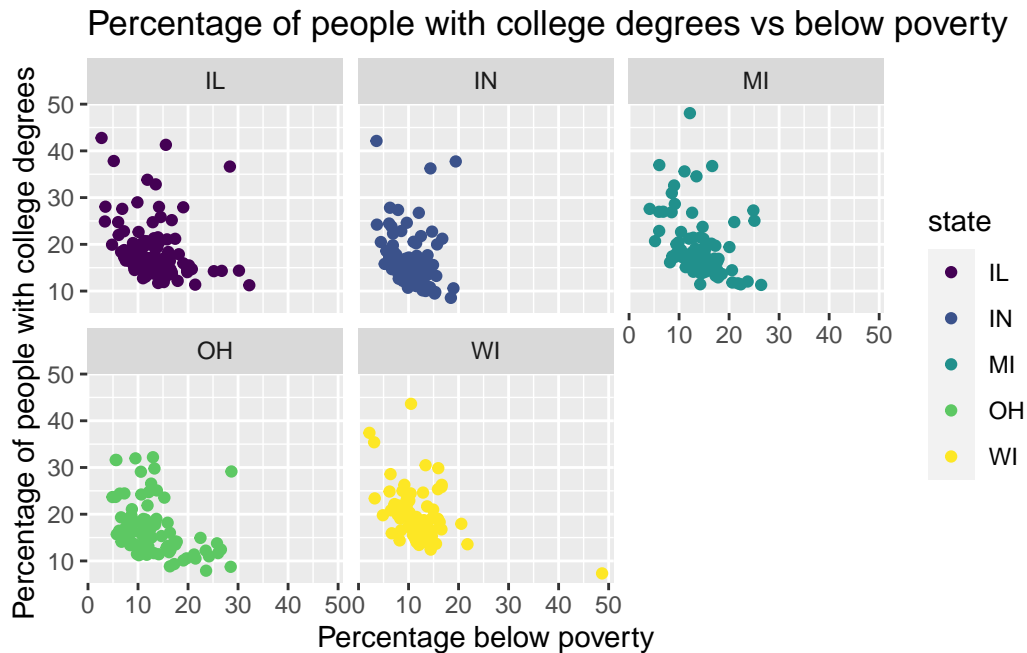
```
ggplot(midwest) +  
  aes(x = popdensity) +  
  geom_histogram(binwidth = 10000) +  
  labs(title = "Population Densities of Midwest Counties", x = "population density", y = "
```



The shape of the distribution is right-skewed. There are some outliers that have a higher population density than the other counties. Most of the counties have a population density between 0 and 25,000 but there a couple of counties in the 60,000 and 80,000 range.

## Exercise 2

```
ggplot(midwest, aes(x = percbelowpoverty, y = percollege, color = state)) +  
  scale_color_viridis_d() +  
  geom_point() +  
  facet_wrap(~state) +  
  labs(title = "Percentage of people with college degrees vs below poverty",  
        x = "Percentage below poverty",  
        y = "Percentage of people with college degrees")
```



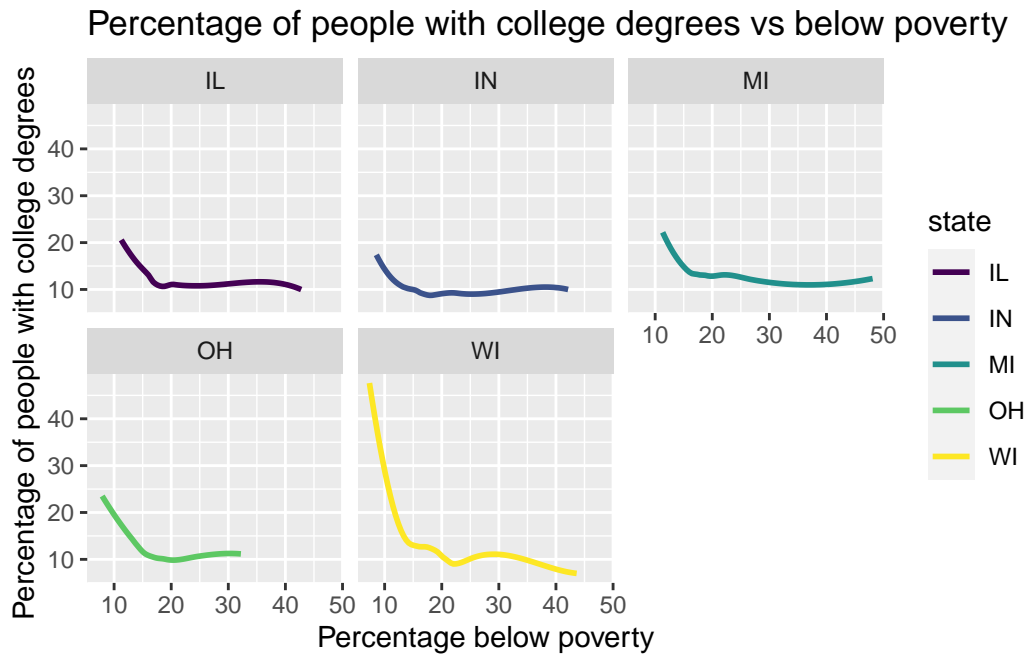
### Exercise 3

There seems to be no correlation between the percentage of people with college degrees and the percentage of people below poverty. However, the plots of the different states are nearly identical. Across all states, the percentage of people below poverty in most counties is 5%-20% and the percentage of people with college degrees is 10%-30%. One difference is that Michigan, Illinois and Ohio have some counties with the percentage of poverty being between 20%-25% as well. There are also different outliers for the different states.

### Exercise 4

```
ggplot(midwest) +
  aes(x = percollege, y = percbelowpoverty, color = state) +
  scale_color_viridis_d() +
  geom_smooth(se = FALSE) +
  facet_wrap(~state) +
  labs(title = "Percentage of people with college degrees vs below poverty",
        x = "Percentage below poverty",
        y = "Percentage of people with college degrees")
```

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



I prefer the plot in Ex. 2 because this plot does not take into account how common certain population densities are in the different midwest states and also does not consider any outliers in the data.

**Exercise 5**

**Exercise 6**

**Exercise 7**