Lab 1 - Data visualization

Alex LaTrenta

Load Packages

```
library(tidyverse)
```

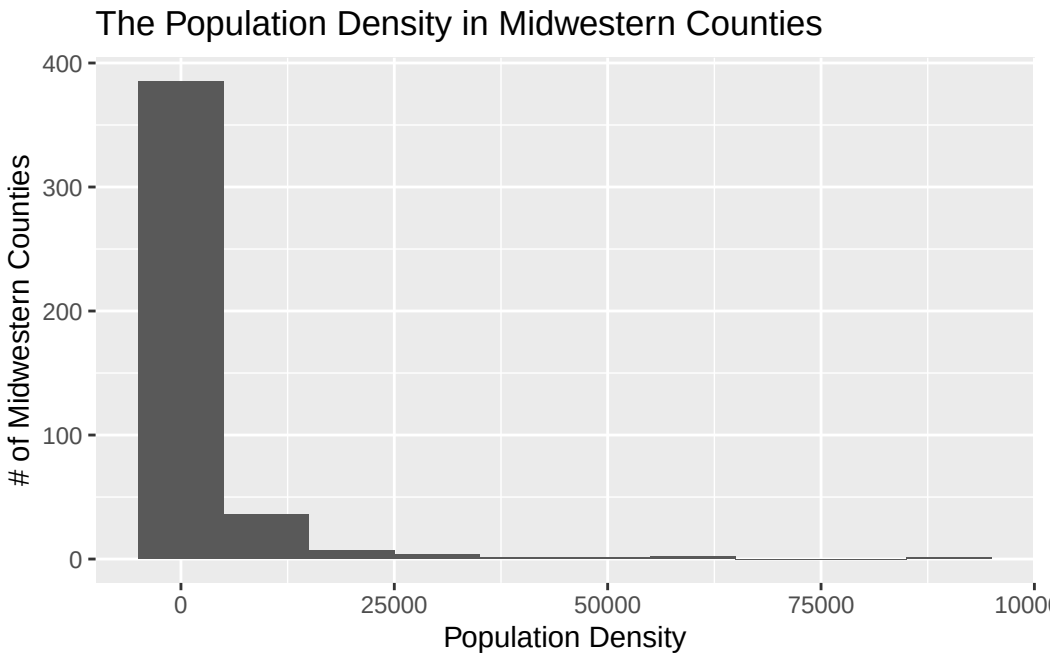
```
Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
had status 1
```

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.)

Problem 1:

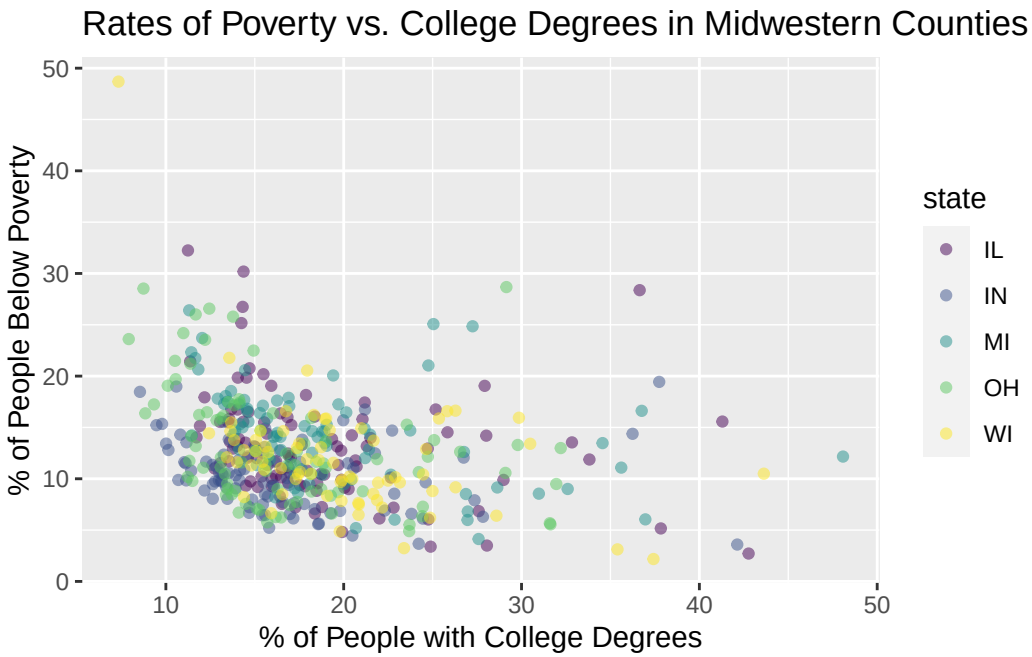
```
ggplot(data = midwest,
       aes(x = popdensity)) +
geom_histogram(binwidth = 10000) +
 labs(x = "Population Density",
       y = "# of Midwestern Counties",
       title = "The Population Density in Midwestern Counties")
```



The graph is very right-skewed with most counties having a population density significantly lower than 25,000 in the low thousands. There are some outliers with medium (~60,000) and large (~85,000) population densities which appear as thin bars on the graph's x axis.

Exercise 2

```
ggplot(data = midwest,
       aes(x = percollege, y = percbelowpoverty, color = state)) +
geom_point(alpha = .5) +
  scale_color_viridis_d() +
 labs(x = "% of People with College Degrees",
       y = "% of People Below Poverty",
       title = "Rates of Poverty vs. College Degrees in Midwestern Counties")
```



Exercise 3

Describe what you observe in the plot from the previous exercise. In your description, include similarities and differences in the patterns across states.

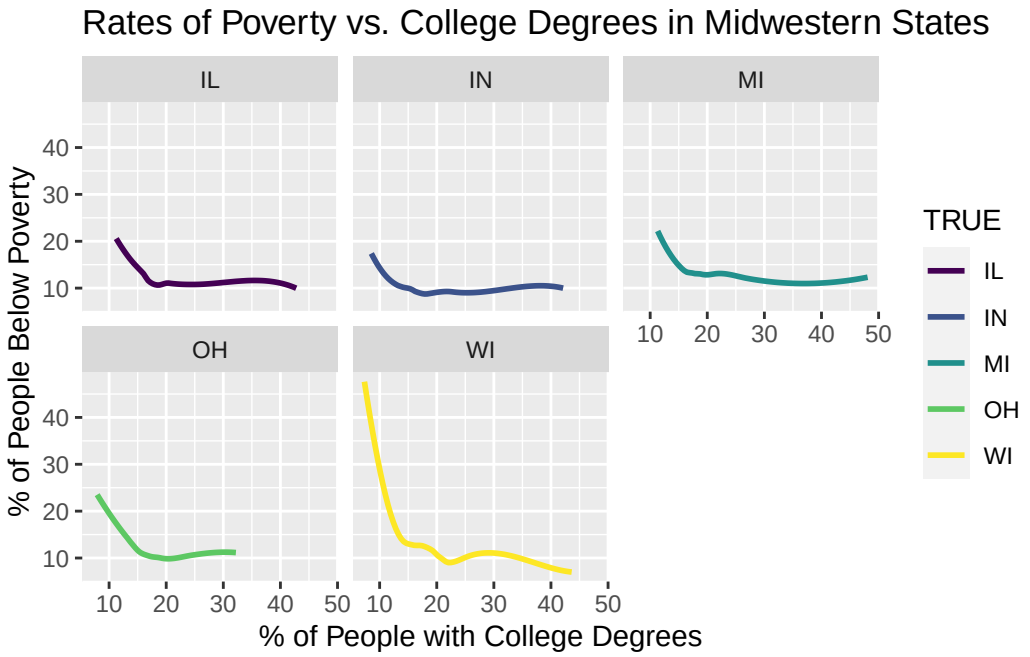
Across most midwestern states, there is a negative linear relationship between the % of people under the poverty line and the % of people with college degrees. This implies that low rates of college degrees may be associated with higher levels of poverty. All states have a large amount of counties with the % College Degrees and % poverty between 10 to 20. OH and MI tend to have outlier counties with low % college degrees and high % poverty despite much variation. WI and IN tend to have more counties with high % college degrees and low % poverty. IL has outlier countries with both of the described characteristics.

Exercise 4

```
ggplot(data = midwest,
       aes(x = percollege, y = percbelowpoverty, color = state)) +
geom_smooth(alpha = .5, se = FALSE) +
  scale_color_viridis_d(TRUE) +
facet_wrap("state") +
 labs(x = "% of People with College Degrees",
```

```
        y = "% of People Below Poverty",
        title = "Rates of Poverty vs. College Degrees in Midwestern States")
```

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



I prefer this plot more because it is easier to visualize the differences between the states when they are separated in a grid. Furthermore, the linear format better summarizes the exact linear trends between the two variables which is visably clearer than a mass of dots. ## Exercise 5

Exercise 6

Exercise 7