Lab 1 - Data visualization

[Ann Chang]

Load Packages

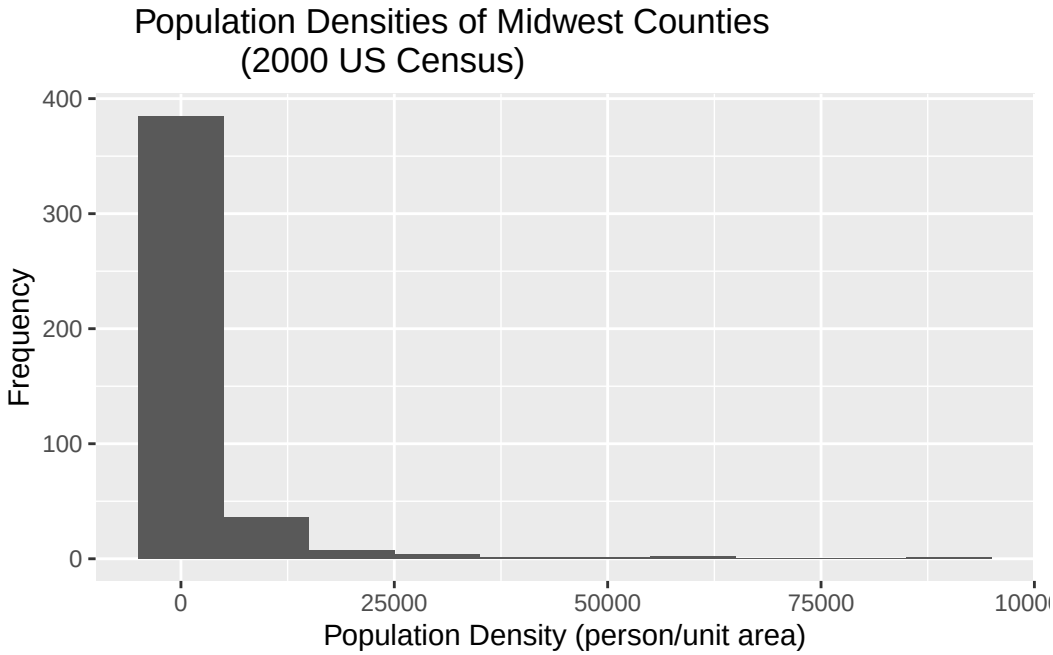
```
library(tidyverse)
```

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

```
library(viridis)
```

Exercise 1

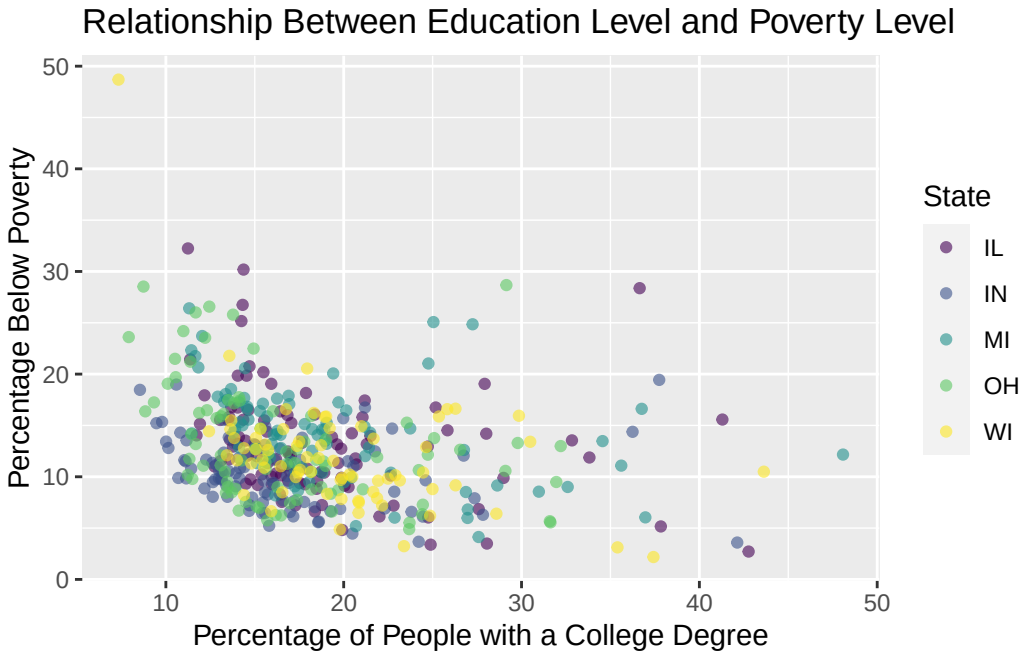
```
ggplot(data = midwest,
       aes(x = popdensity)) +
  geom_histogram(binwidth = 10000) +
  labs(
    x = "Population Density (person/unit area)",
    y = "Frequency",
    title =
      "    Population Densities of Midwest Counties
              (2000 US Census)"
    )
```



- The distribution of the data seem to be right-skewed.
- There seem to be some outliers of relatively large population densities, such as the incidences beyond 50000 people/unit area.

Exercise 2

```
ggplot(data = midwest,
       aes(x = percollege,
           y = percbelowpoverty,
           color = state)) +
  geom_point(alpha = 0.6) +
  labs(
    x = "Percentage of People with a College Degree",
    y = "Percentage Below Poverty",
    title = "Relationship Between Education Level and Poverty Level",
    color = "State"
  ) +
  scale_color_viridis_d()
```



Exercise 3

There seems to be a negative correlation between the percentage of people below the poverty line and the percentage of people with a college degree. Illinois seems to have relatively more outlying counties where the percentage of below poverty individuals is larger than what would be expected from the percentage of people with a college degree. This is also seen in some counties in Ohio. The counties in Wisconsin are relatively clustered closely together, with no counties' percentage of individuals below poverty exceeding above 23%.

Exercise 4

Exercise 5

Exercise 6

Exercise 7