# Lab 1 - Data visualization

## Ayden Frost

## Questions

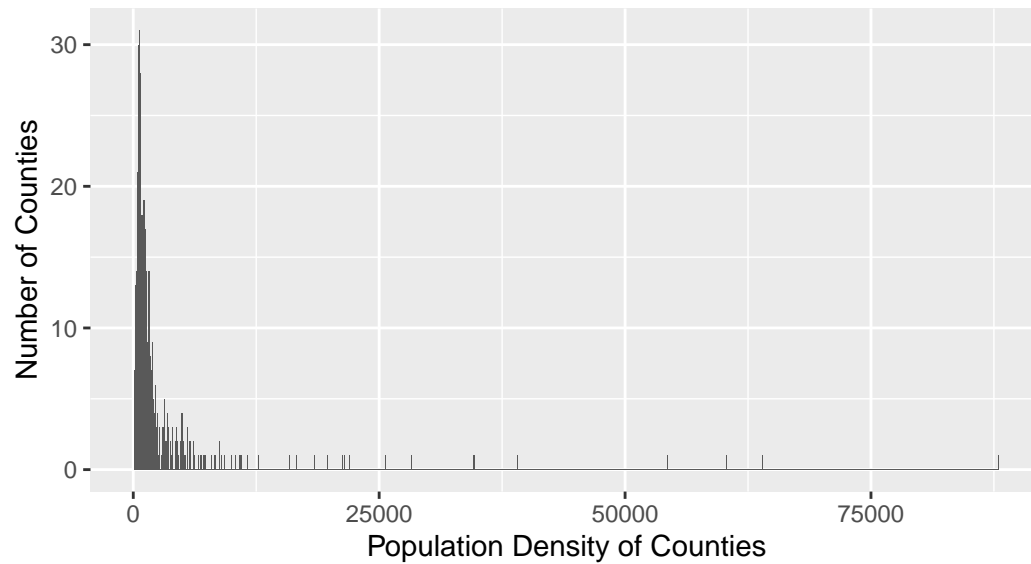### Part 1

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.6
v forcats   1.0.1     v stringr   1.6.0
v ggplot2   4.0.1     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.2
v purrr     1.2.0
-- Conflicts ------------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becon
```

### Question 1

```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 100) +
  labs(
    x = "Population Density of Counties",
    y = "Number of Counties",
    title = "Population Density of Midwesten Counties",
    subtitle = "Binwidth = 100"
  )
```

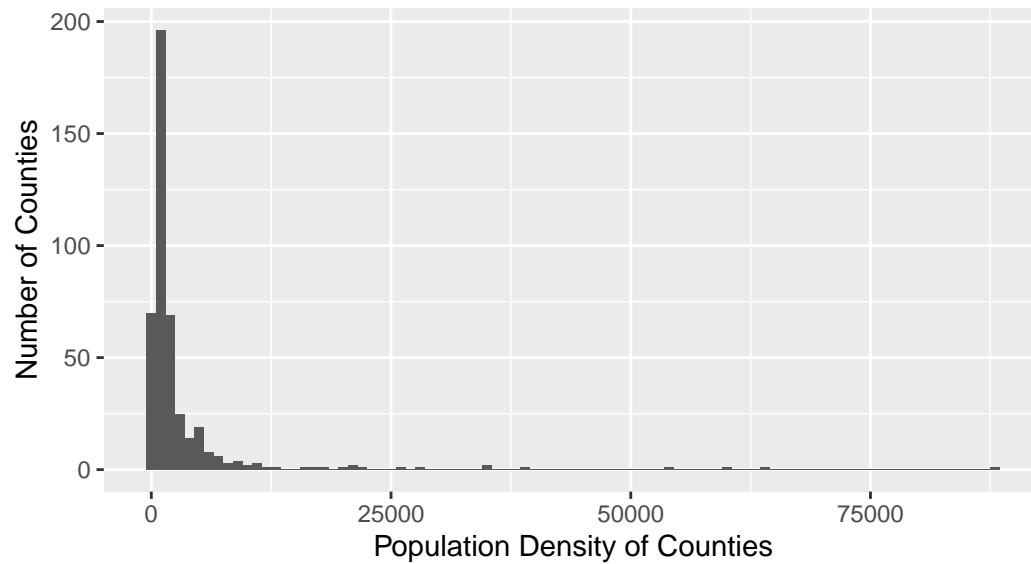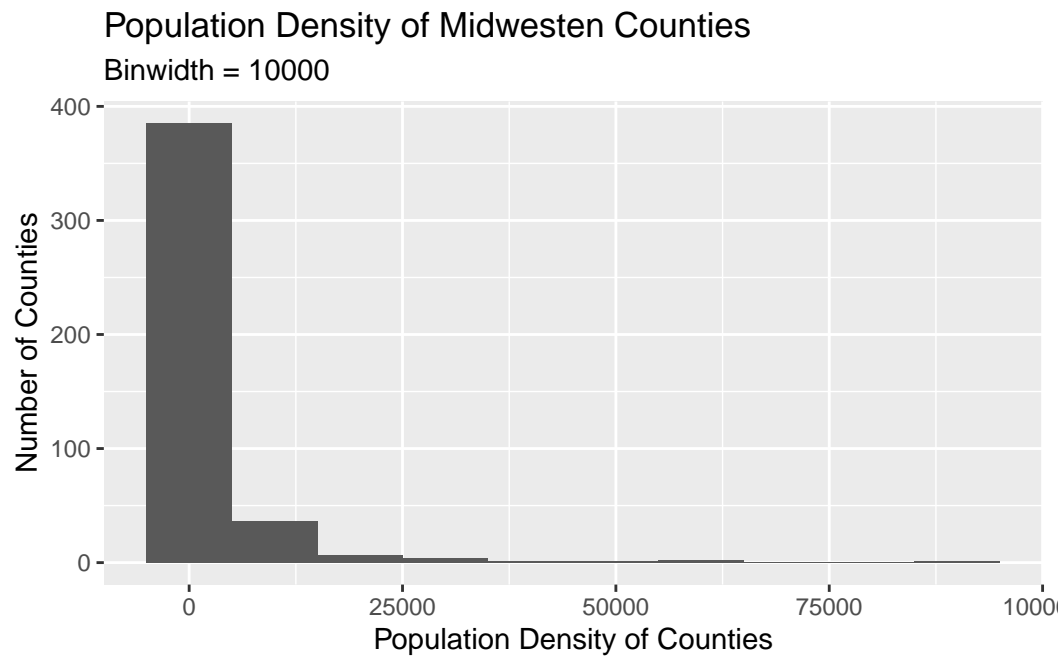## Population Density of Midwesten Counties
Binwidth = 100



```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 1000) +
  labs(
    x = "Population Density of Counties",
    y = "Number of Counties",
    title = "Population Density of Midwesten Counties",
    subtitle = "Binwidth = 1000"
  )
```

## Population Density of Midwesten Counties
Binwidth = 1000



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

3

## Population Density of Midwesten Counties
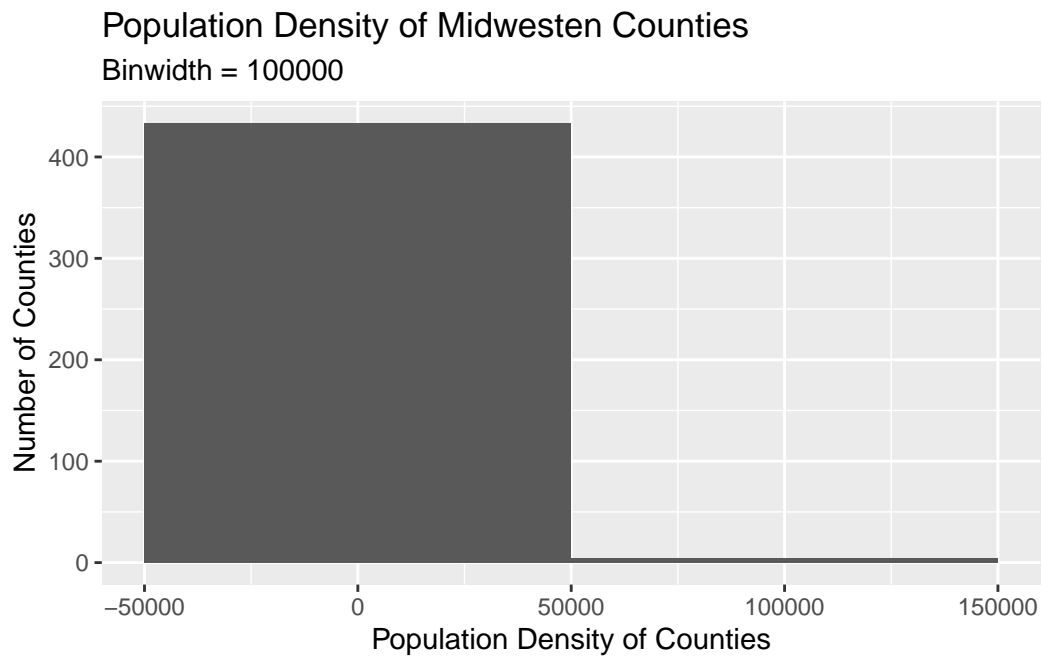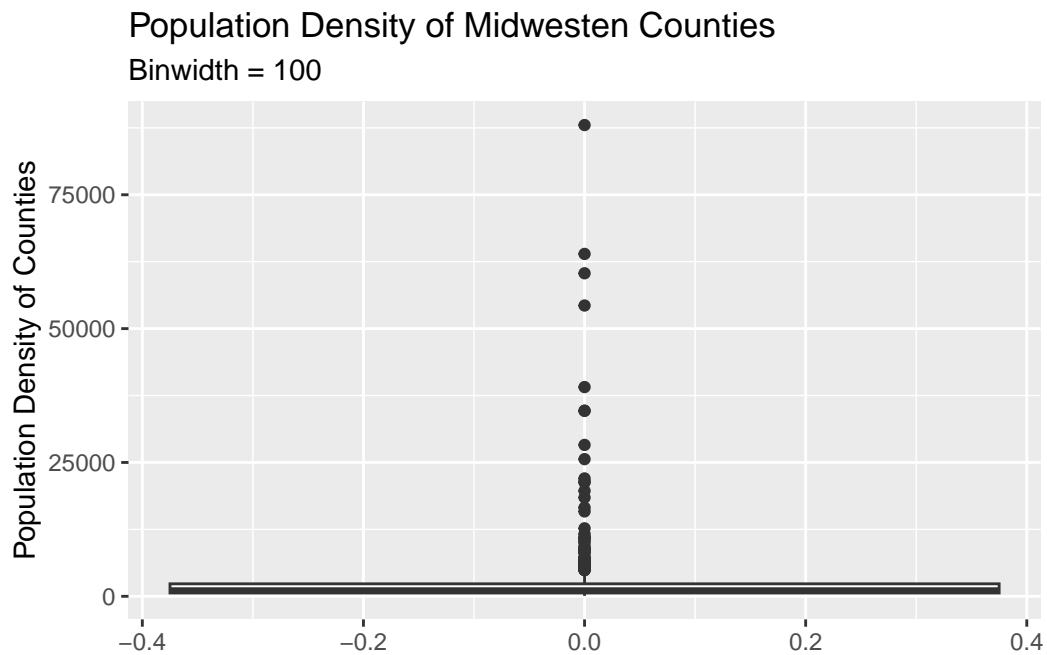Binwidth = 10000



```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 100000) +
  labs(
    x = "Population Density of Counties",
    y = "Number of Counties",
    title = "Population Density of Midwesten Counties",
    subtitle = "Binwidth = 100000"
  )
```

## Population Density of Midwesten Counties
Binwidth = 100000



The 1000 Binwidth histogram would be ideal as it visualizes the data in a presentable way that is easy to interpret when compared to the other 3 histograms.
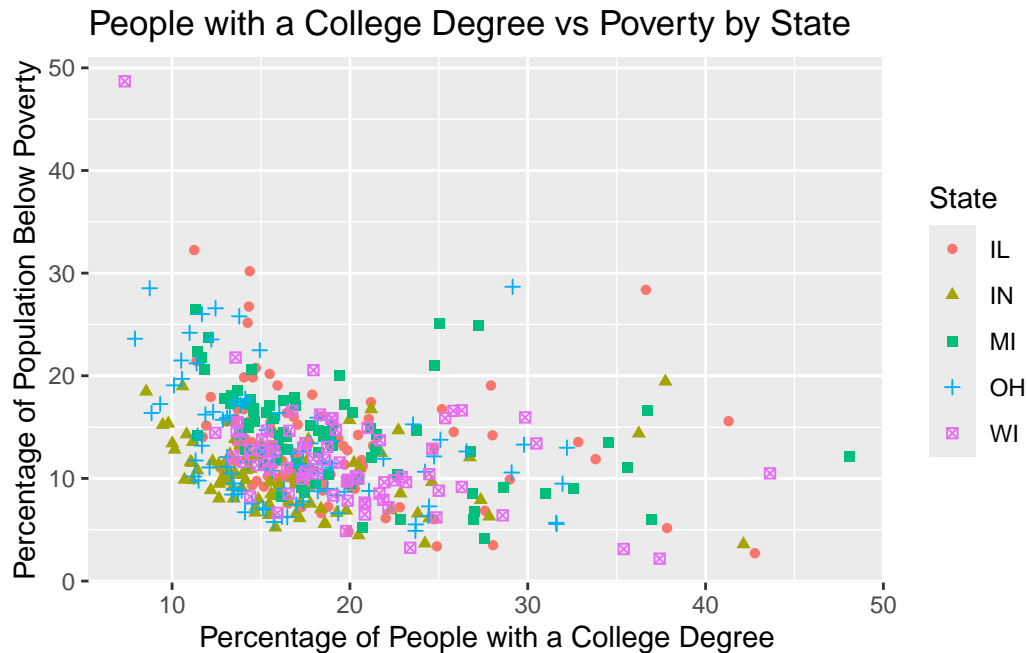
**Question 2**

```
ggplot(midwest, aes(y = popdensity)) +
  geom_boxplot() +
  labs(
      y = "Population Density of Counties",
    title = "Population Density of Midwesten Counties",
    subtitle = "Binwidth = 100"
  )
```

## Population Density of Midwesten Counties
Binwidth = 100



The distribution of population sizes amongst counties displayed in both the histogram and boxplot show that most counties tend to be under 12,500 individuals per unit area, with only a small handful of counties breaking the threshold. One county that stands out is Cook county, as it has a population density of roughly 88,000. The most likely reason behind this could be the existence of a city or larger community within the county.

### Question 3

```
ggplot(midwest, aes(y = percbelowpoverty, x = percollege, color = state, shape = state)) +
  geom_point() +
  labs(
    color = "State",
    shape = "State",
    x = "Percentage of People with a College Degree",
    y = "Percentage of Population Below Poverty",
    title = "People with a College Degree vs Poverty by State"
  )
```
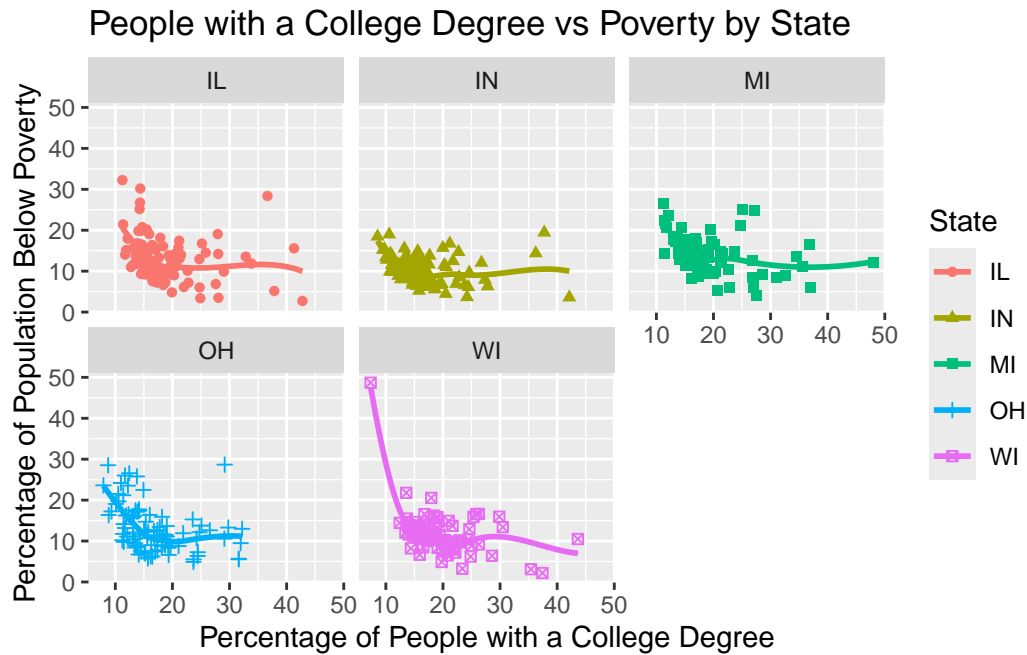
The scatterplot shows that as populations become more educated and have degrees, they tend to see fewer people in poverty when compared to less educated populations. The one county that stood out greatly is MENOMINEE county where nearly 50% of the population sits below the poverty line while only roughly 7% have a degree. It would be very difficult to determine the state trends as the data within the scatterplot is very congested and hard to read if trying to make inferences on the states.

**Question 4**

```
ggplot(midwest, aes(y = percbelowpoverty, x = percollege, color = state, shape = state)) +
  geom_point() +
  facet_wrap(~ state) +
  geom_smooth(se = FALSE) +
  labs(
    color = "State",
    shape = "State",
    x = "Percentage of People with a College Degree",
    y = "Percentage of Population Below Poverty",
    title = "People with a College Degree vs Poverty by State"
  )
```

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

## People with a College Degree vs Poverty by State



This new plot definitely makes analyzing the data by state much easier when compared to the heavily congested scatterplot in question 3. The data is much more clear and concise in the facet wrapped scatter plot.

**Question 5**

**Question 6**

**Question 7**

**Part 2**

**Enough about the Midwest!**

**Question 8**

**Question 9**