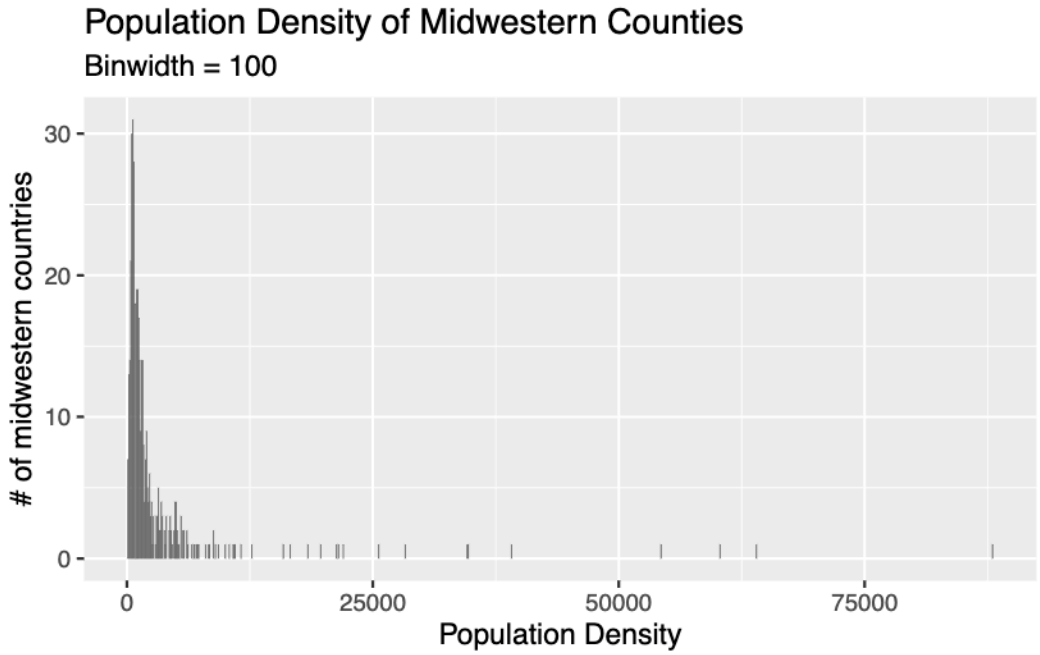
Lab 2 - Data Visualization

Brandon Leslie

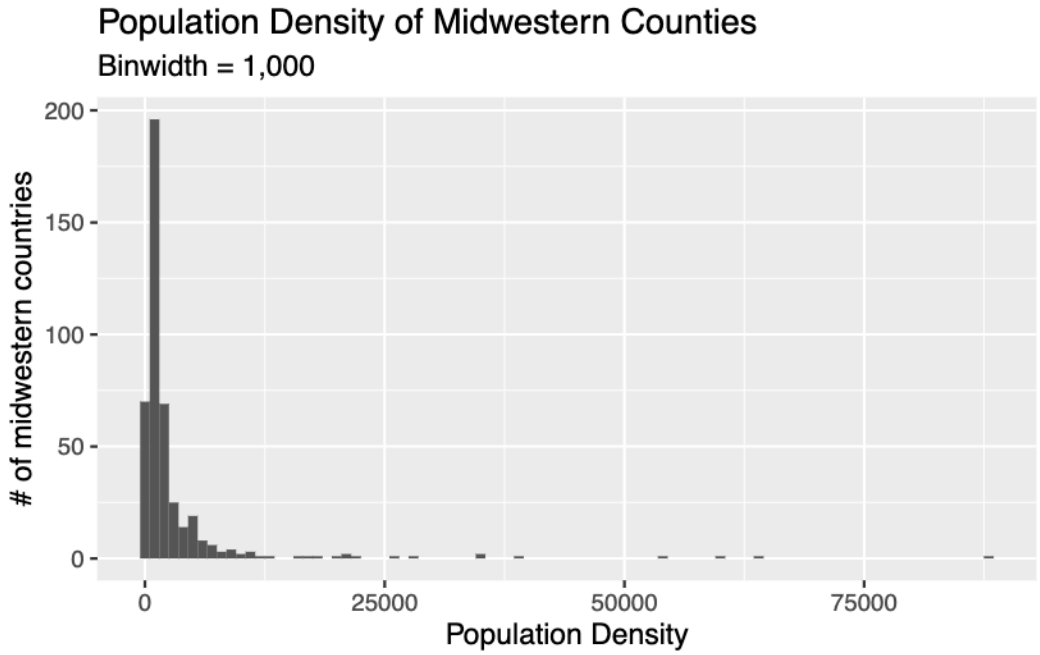
Part 1

Question 1

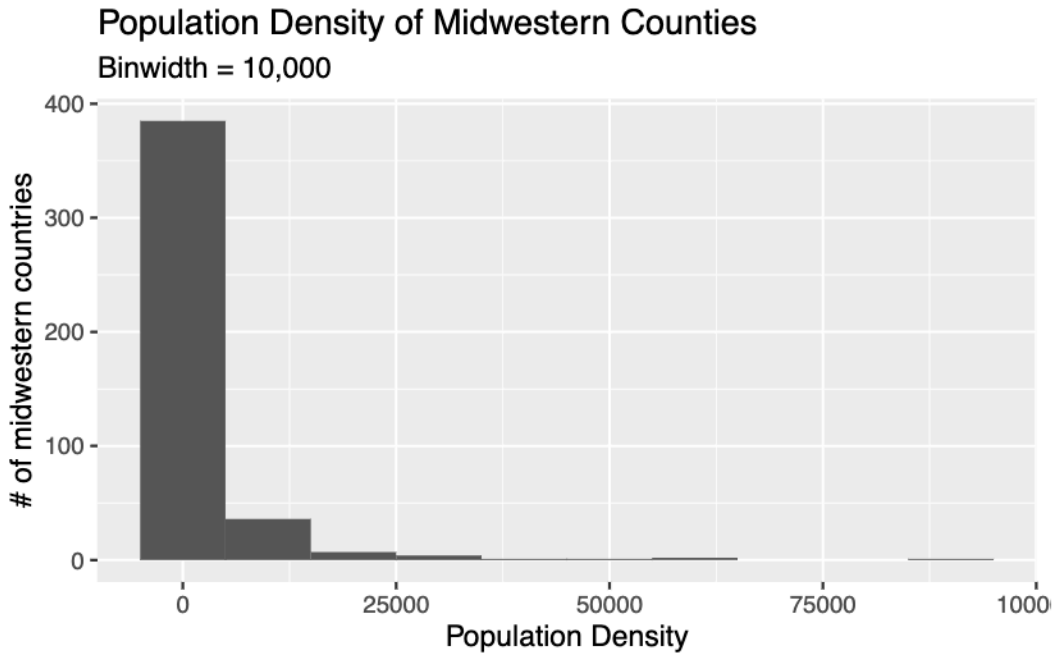
Binwidth = 100



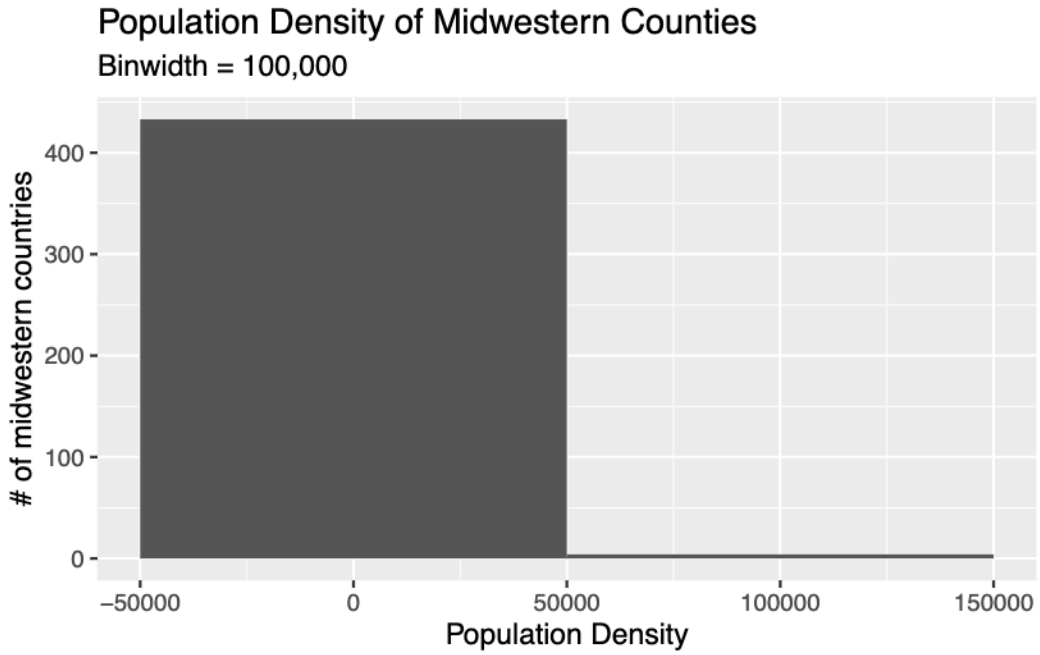
Binwidth = 1,000



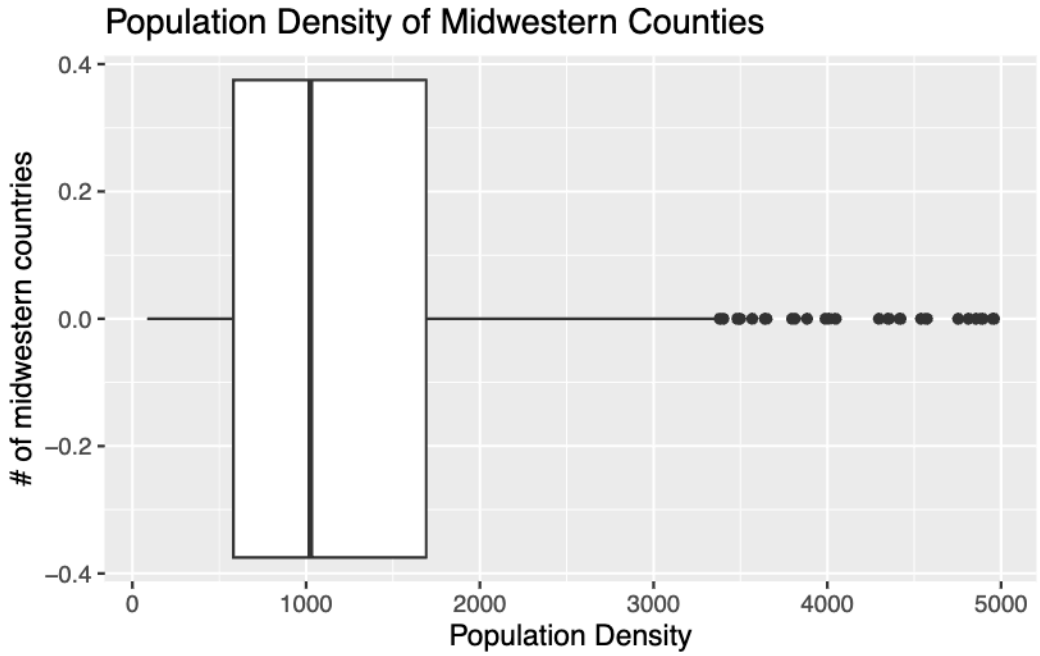
Binwidth = 10,000



Binwidth = 100,000



Question 2

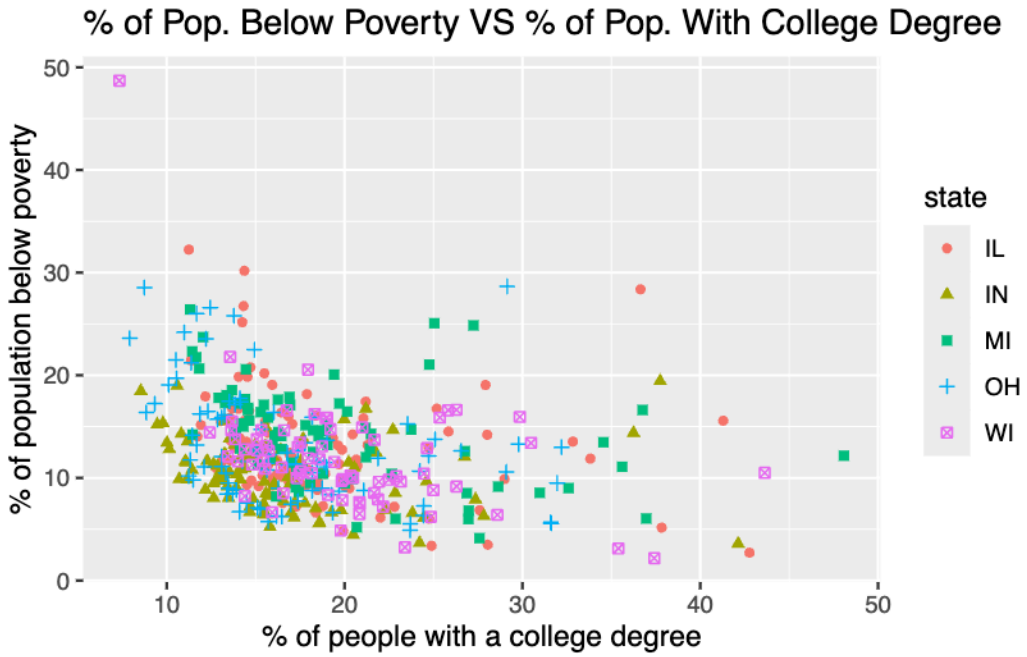


The population density for the Midwestern counties isn't best visualized by a box plot. However, we can draw some conclusions with this graphs. For instance, median population of this data set is around the **1,000's** mark. However, the mean is close to the **3,000's**. This is due to the heavy presence of outliers to the right of the boxplot. One outlier from the data-set is **COOK** county (as seen below).

```
# A tibble: 1 x 3
  popdensity county     n
       <dbl> <chr>  <int>
1     88018. COOK       1
```

COOK county is an outlier because of the fact that it is so far away from not just the mean, and the median. The *popdensity* (population density) of **COOK** county is more than **20x** the mean, and is considerably far from the median.

Question 3

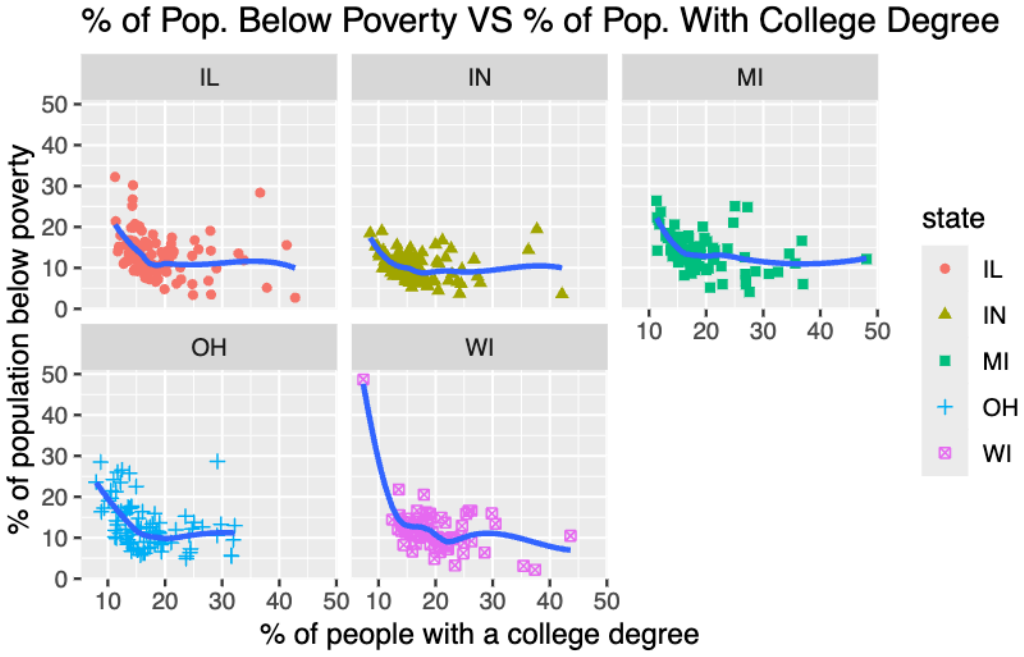


The overall relationship between % of people with a college degree and % below poverty is mostly scattered. However, the trend of the graph seems to be going down. By what we can tell, as % of people with a college degree increases, poverty decreases, and as poverty increases, % of people with a college degree decreases.

```
# A tibble: 1 x 3
  percbelowpoverty county        n
             <dbl> <chr>     <int>
1             48.7 MENOMINEE     1
```

MENOMINEE county is an outlier because it is the only county to be above the 40% line, with no other county being anywhere close to that certain percentage.

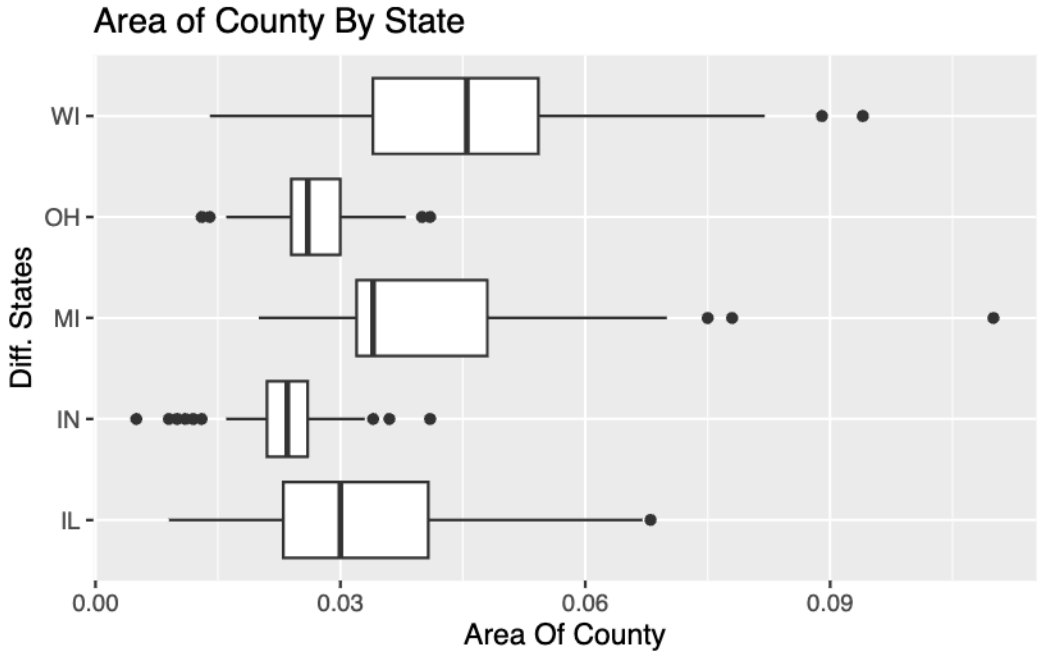
Question 4



I prefer this plot more because it's easier to visualize and compare each state. The multiple different colors and shapes in the same graph make it hard to draw conclusions about the states, making it harder to analyze the the graph. However, when separated, individual conclusions can be compared more clearly.

Question 5

Do some states have counties that tend to be geographically larger than others?

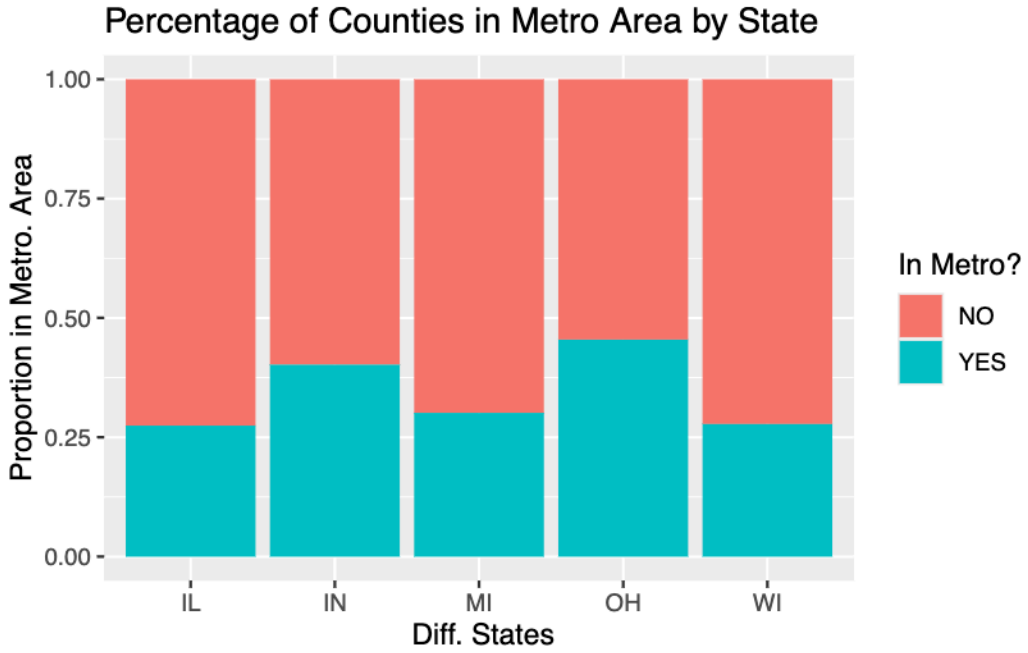


Typical county area sizes are bigger in states like **Mississippi**, **Wisconsin**, and **Illinois**. The variability of county sizes across each state are higher in those with larger county sizes then smaller. The state with the single largest county is **Mississippi**. The name of this county is...

```
# A tibble: 1 x 3
   area county        n
  <dbl> <chr>     <int>
1  0.11 MARQUETTE     1
```

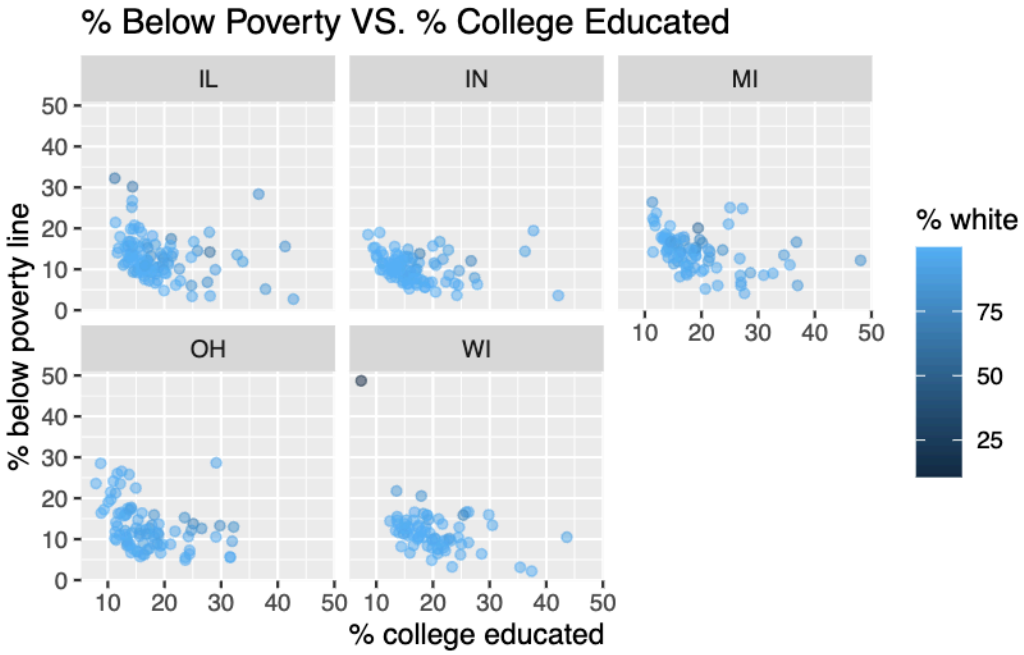
Question 6

Do some states have a higher percentage of their counties located in a metropolitan area?



Some states **DO** have a higher percentage of their counties located in metropolitan area. States like **Indiana** and **Ohio** have around 40% of their counties in metropolitan areas, while other states like **Illinois**, **Mississippi**, and **Wisconsin** have around 25%-30% of their counties in metropolitan areas.

Question 7



The one clear outlier in **Wisconsin** is...

```
# A tibble: 1 x 3
  county    percbelowpoverty percollege
  <chr>                <dbl>      <dbl>
1 MENOMINEE             48.7       7.34
```

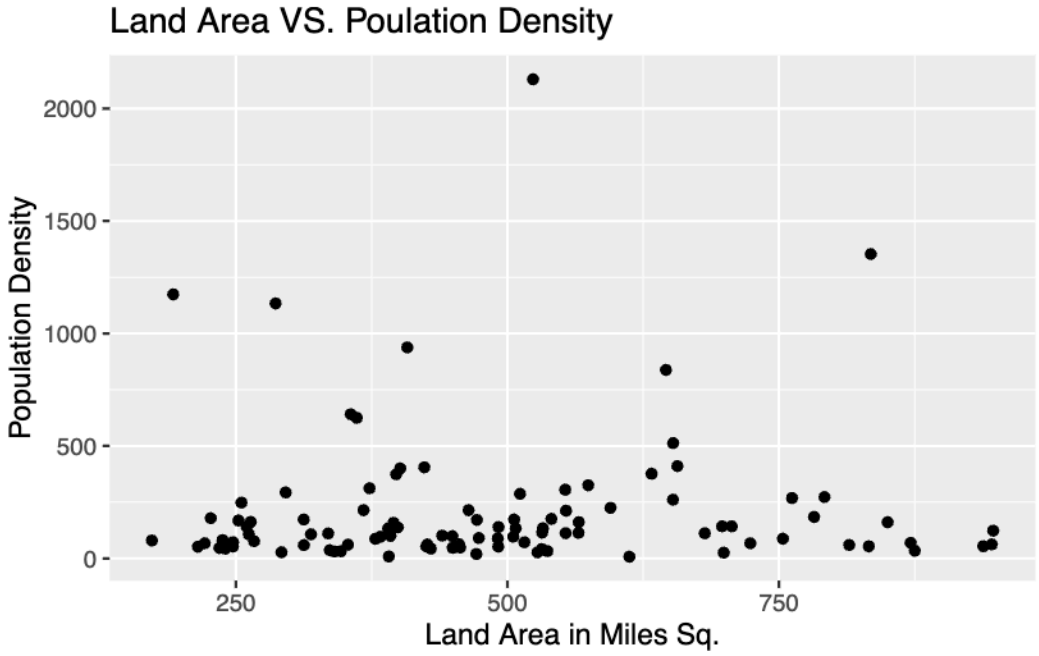
MENOMINEE county has a high density of *popamerindian* that reside there, with an **89%** chance of dominance. There is little to no population with other races, with *popwhite* being the next leading race, consisting of **10%**, and *popothers* with less than **1%**.

Part 2

Question 8

Guess the relationship between *popdensity & area*

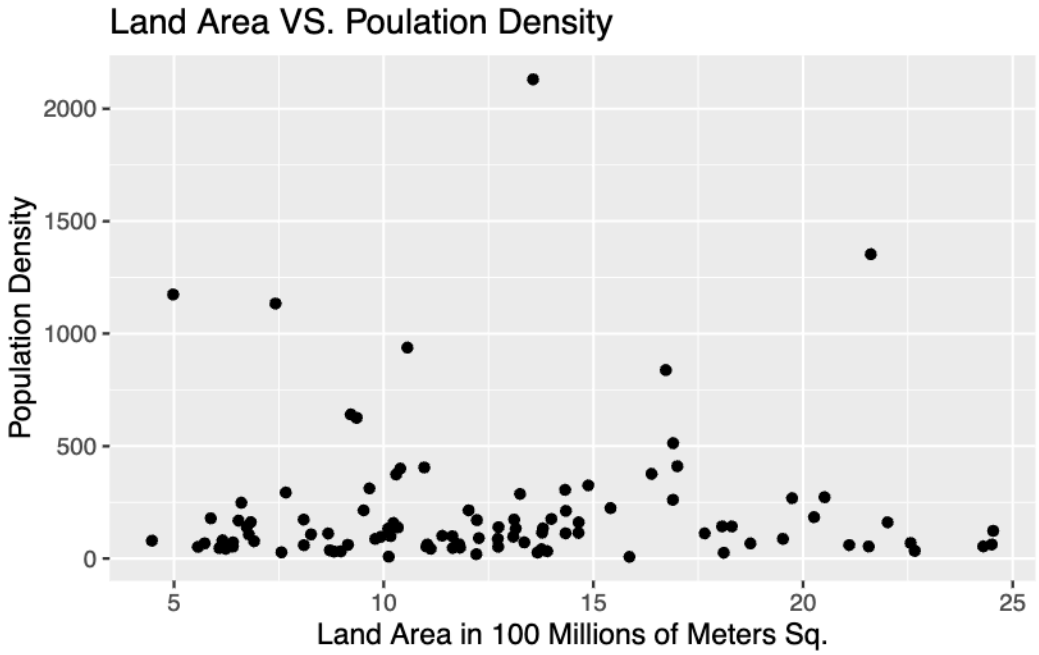
I think relationship between land area & population density will be an inverse relationship. (Negative)



Conclusion:

The relationship between popdensity and area is non-existant. From a humans persepctive, there seems ot be no relationship between

Question 9



There is relatively no change to the way that this data is formatted from question 8. This is due to the fact that changing the x-value proportionately will have no change to the y value, as long as the y-value is not directly dependent to the x value. In this case, Miles Squared have been proportionately transformed to meters squared, and they both equate to the same value when converted. So the scatter plot will be visually identical in position of the graph to both the computer, and the human eye, when scaled correctly.