Lab 1 - Data visualization

Ayden Frost

Questions

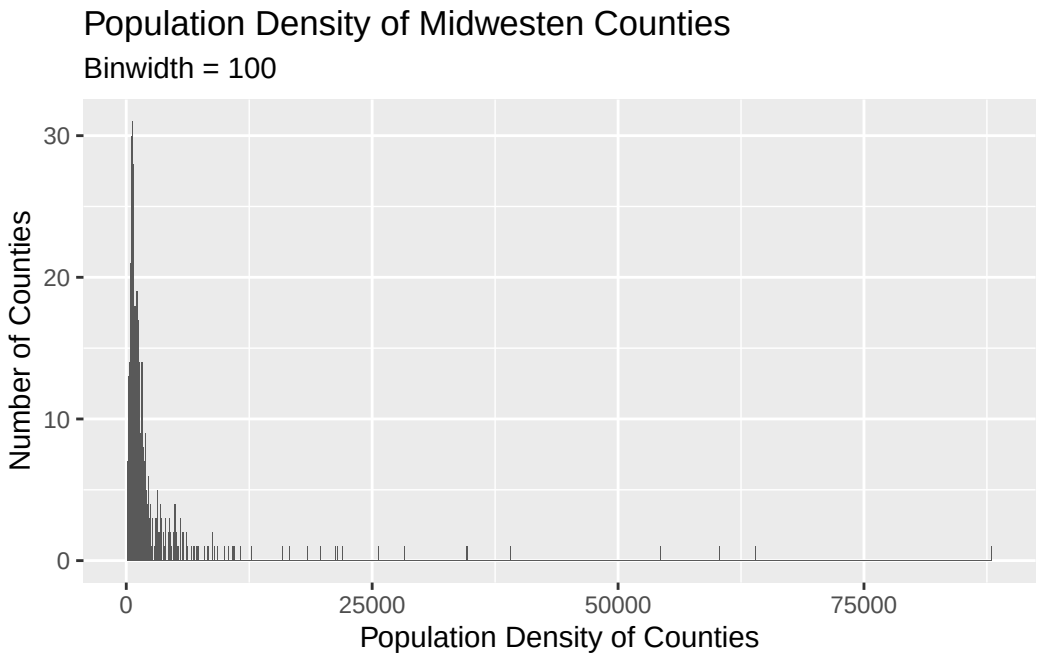
Part 1

```
library(tidyverse)
```

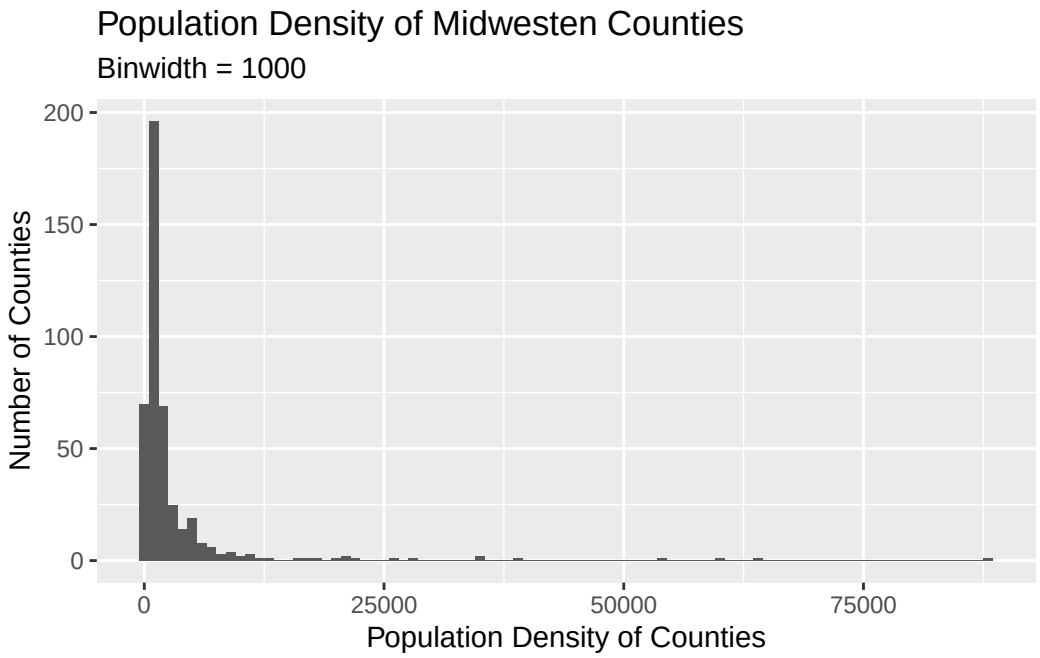
```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.6
v forcats   1.0.1     v stringr   1.6.0
v ggplot2   4.0.1     v tibble    3.3.0
v lubridate 1.9.4     v tidyr     1.3.2
v purrr     1.2.0
-- Conflicts ------------------------------------------ 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 beco
```

Question 1

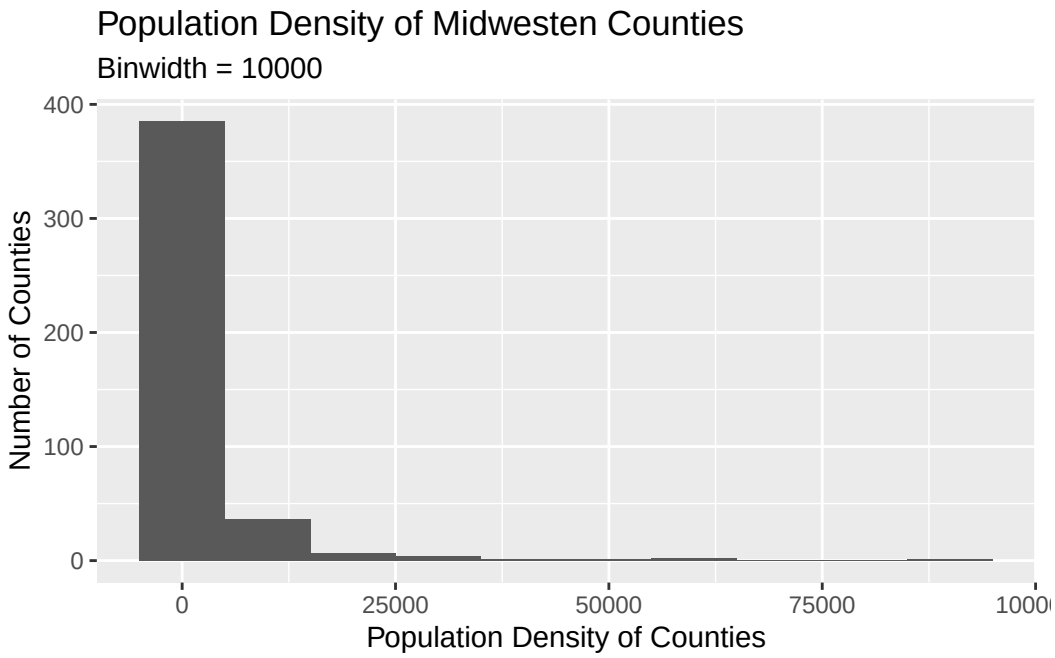
```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 100) +
  labs(
    x = "Population Density of Counties",
    y = "Number of Counties",
    title = "Population Density of Midwesten Counties",
    subtitle = "Binwidth = 100"
  )
```



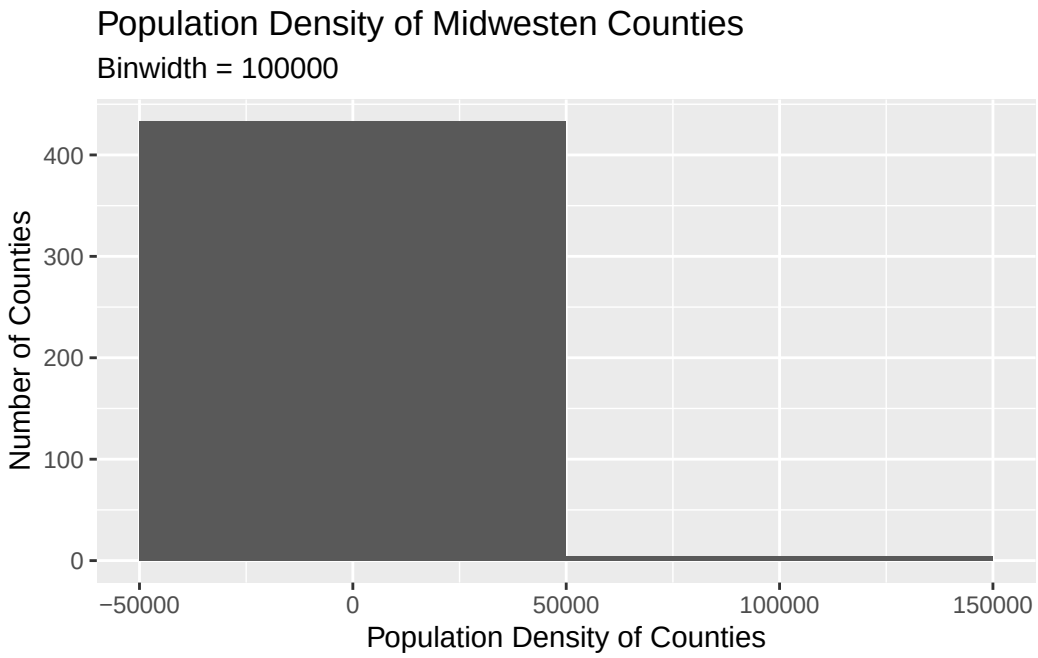

```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 1000) +
  labs(
    x = "Population Density of Counties",
    y = "Number of Counties",
    title = "Population Density of Midwesten Counties",
    subtitle = "Binwidth = 1000"
  )
```




```
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"
  )
```



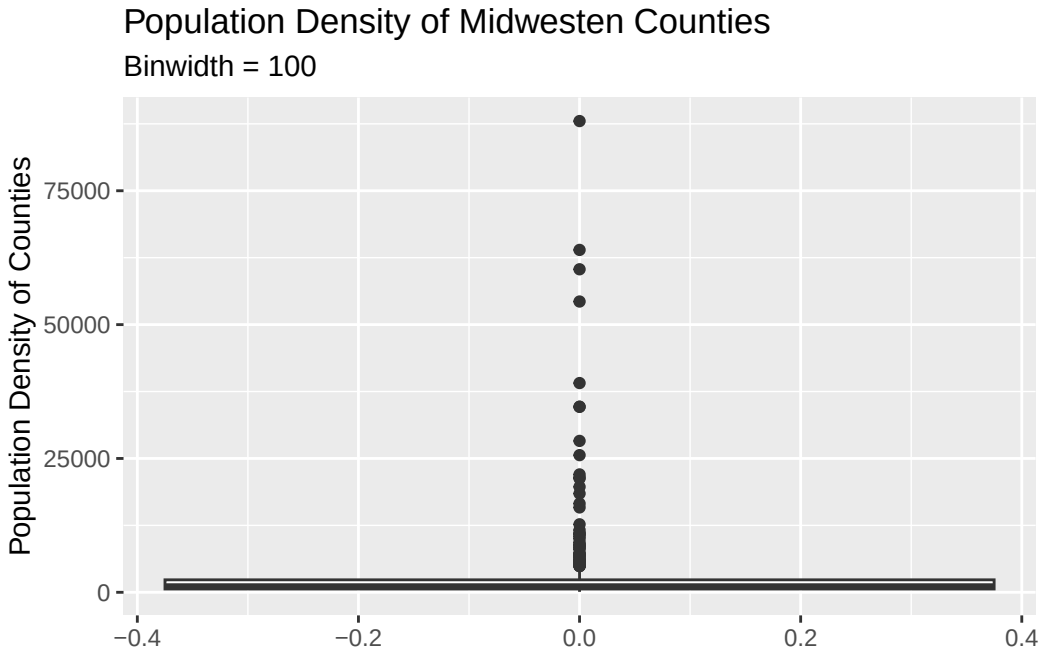

```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 100000) +
  labs(
    x = "Population Density of Counties",
    y = "Number of Counties",
    title = "Population Density of Midwesten Counties",
    subtitle = "Binwidth = 100000"
  )
```



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"
  )
```



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

Question 4

Question 5

Question 6

Question 7

Part 2

Enough about the Midwest!

Question 8

Question 9