Lab 1 - Data visualization

Deandra Rasheesa Maheswari

Questions

Part 1

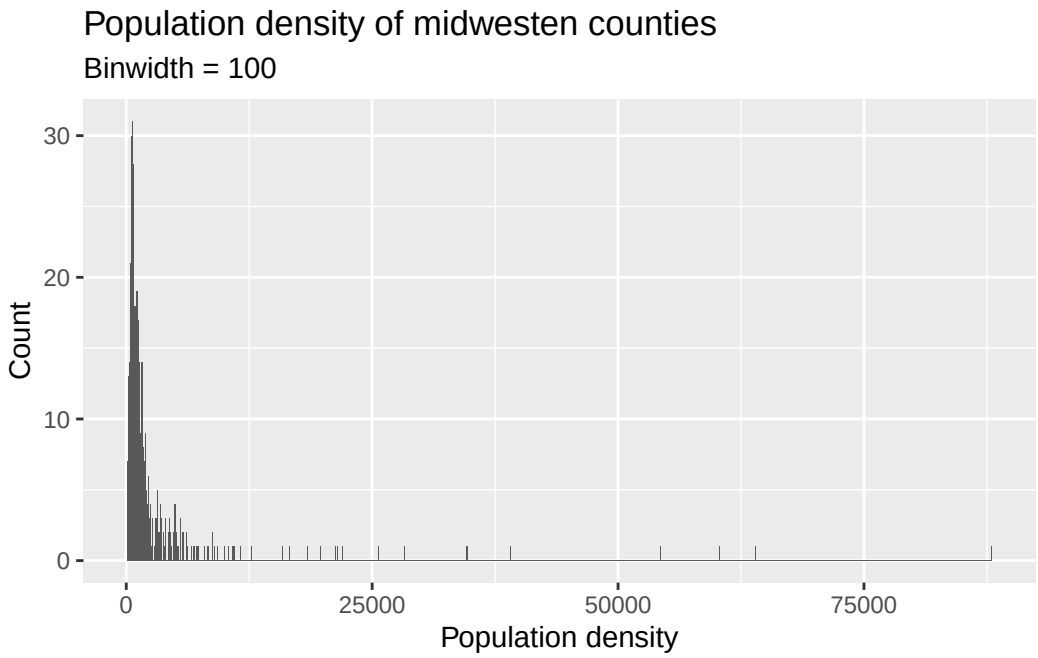
```
library(tidyverse)
```

```
-- 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 becor
```

Question 1

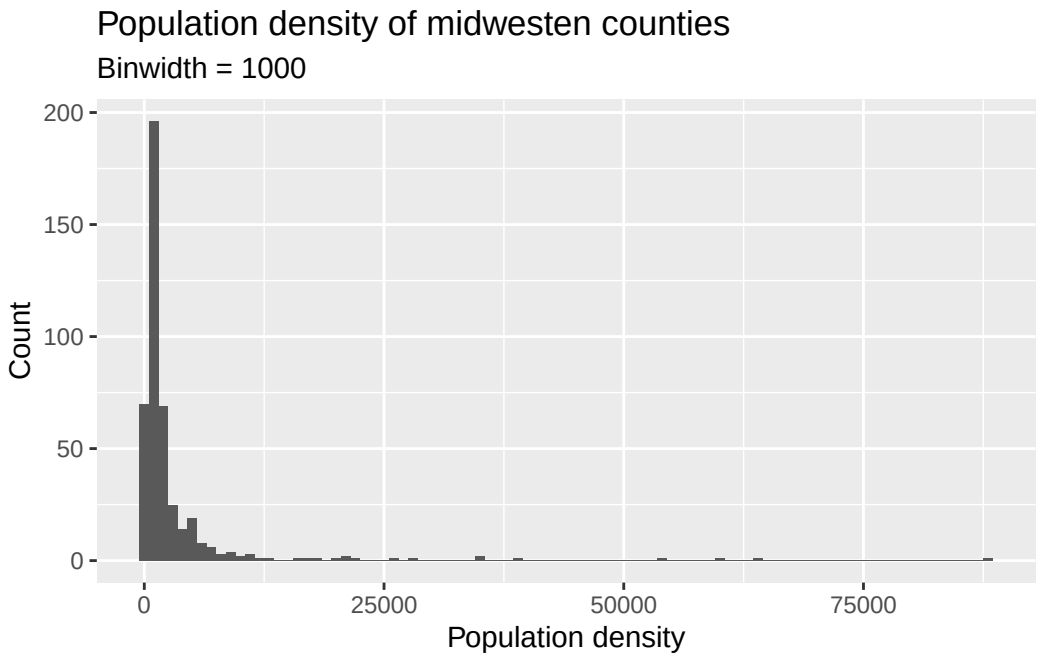
Binwidth = 100

```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 100) +
  labs(
    x = "Population density",
    y = "Count",
    title = "Population density of midwesten counties",
    subtitle = "Binwidth = 100"
  )
```



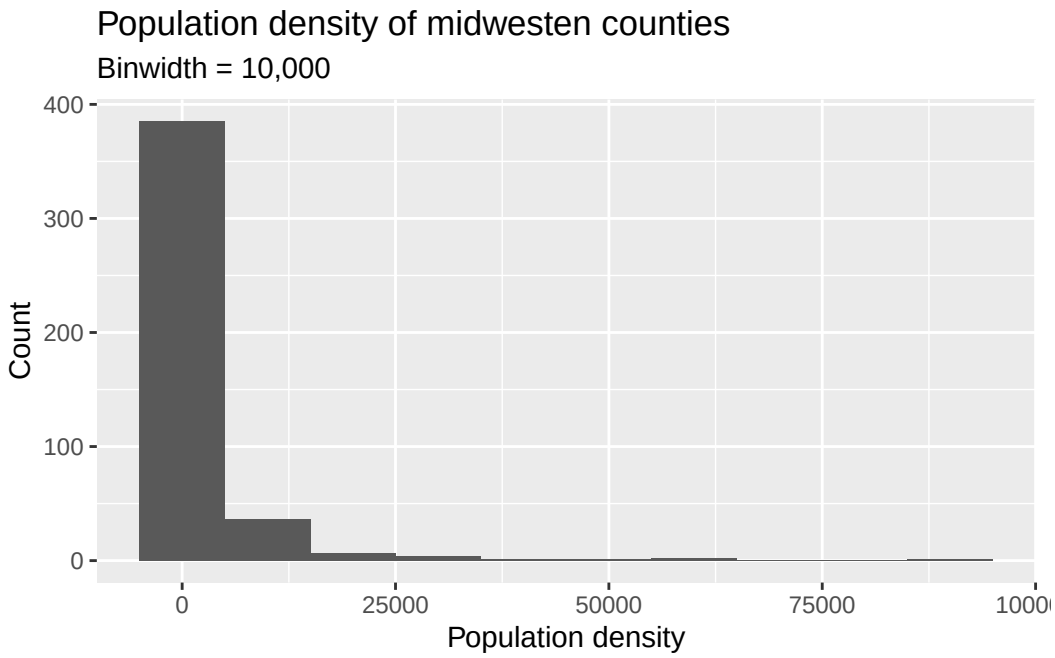
Binwidth = 1000

```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 1000) +
  labs(
    x = "Population density",
    y = "Count",
    title = "Population density of midwesten counties",
    subtitle = "Binwidth = 1000"
  )
```



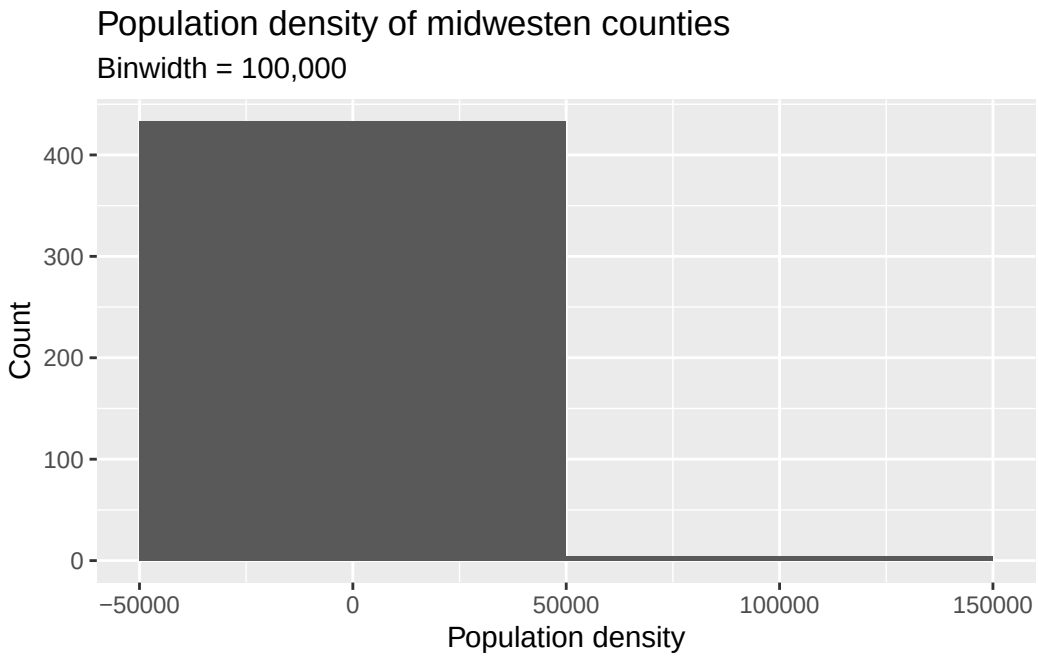
Binwidth = 10,000

```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 10000) +
  labs(
    x = "Population density",
    y = "Count",
    title = "Population density of midwesten counties",
    subtitle = "Binwidth = 10,000"
  )
```



Binwdith = 100,000

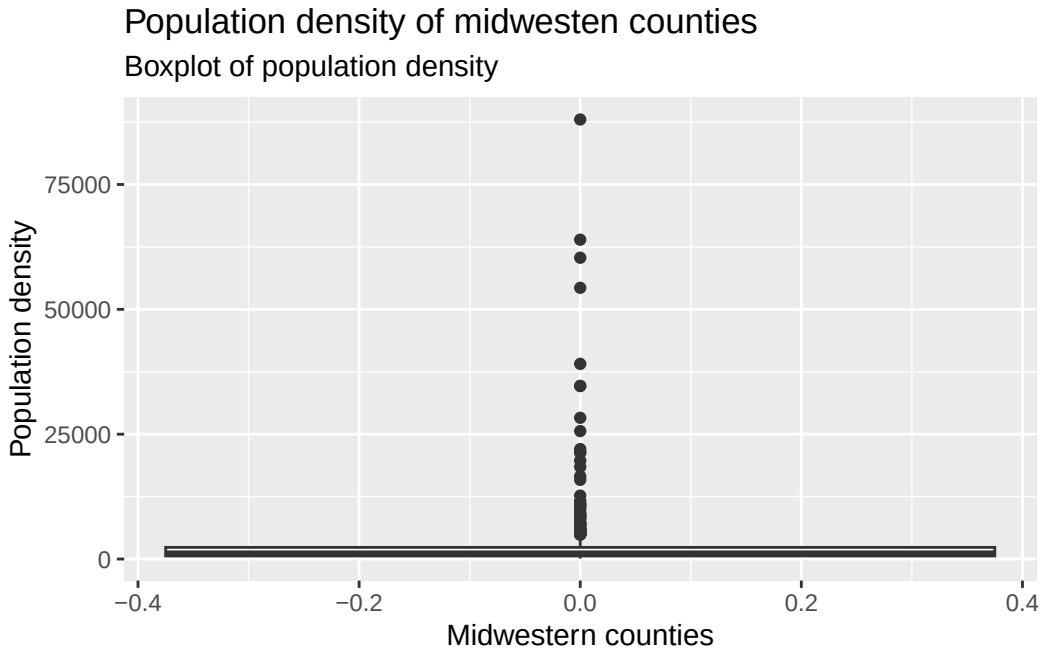
```
ggplot(midwest, aes(x = popdensity)) +
  geom_histogram(binwidth = 100000) +
  labs(
    x = "Population density",
    y = "Count",
    title = "Population density of midwesten counties",
    subtitle = "Binwidth = 100,000"
  )
```



Comment: A graph with banwidth 1,000 is the most appropriate because it shows the overall distribution clearly without losing important detail, not too noisy nor oversmooth.

Question 2

```
ggplot(midwest, aes(y = popdensity)) +
  geom_boxplot() +
  labs(
    x = "Midwestern counties",
    y = "Population density",
    title = "Population density of midwesten counties",
    subtitle = "Boxplot of population density"
  )
```




```
view(midwest)
```

Comment: The distribution of population density among Midwestern counties is highly right-skewed, which means most counties having relatively low population density. One clear outlier is Cook County, IL, which has much higher population density than most other counties.

Question 3

Question 4

Question 5

Question 6

Question 7

Part 2

Enough about the Midwest!

Question 8

Question 9