lab 1

Claudia Carugati

Quarto

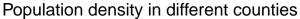
Running Code

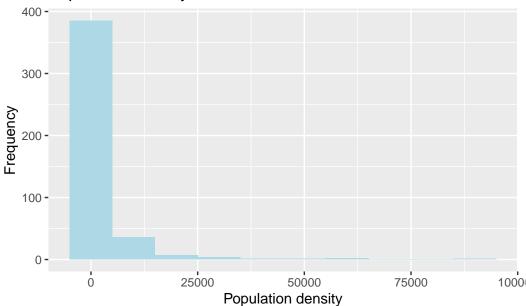
```
library(tidyverse)
Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
had status 1
-- Attaching packages ----- tidyverse 1.3.2 --
v ggplot2 3.3.6 v purrr
                         0.3.4
v tibble 3.1.8 v dplyr
                       1.0.9
v tidyr 1.2.0 v stringr 1.4.1
v readr
        2.1.2
                v forcats 0.5.2
-- Conflicts ------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag() masks stats::lag()
  library(viridis)
Loading required package: viridisLite
  glimpse(midwest)
```

```
Rows: 437
Columns: 28
$ PID
                       <int> 561, 562, 563, 564, 565, 566, 567, 568, 569, 570,~
                       <chr> "ADAMS", "ALEXANDER", "BOND", "BOONE", "BROWN", "~
$ county
                       <chr> "IL", "IL", "IL", "IL", "IL", "IL", "IL", "IL", "IL", "~
$ state
                       <dbl> 0.052, 0.014, 0.022, 0.017, 0.018, 0.050, 0.017, ~
$ area
$ poptotal
                       <int> 66090, 10626, 14991, 30806, 5836, 35688, 5322, 16~
$ popdensity
                       <dbl> 1270.9615, 759.0000, 681.4091, 1812.1176, 324.222~
                       <int> 63917, 7054, 14477, 29344, 5264, 35157, 5298, 165~
$ popwhite
$ popblack
                       <int> 1702, 3496, 429, 127, 547, 50, 1, 111, 16, 16559,~
                       <int> 98, 19, 35, 46, 14, 65, 8, 30, 8, 331, 51, 26, 17~
$ popamerindian
                       <int> 249, 48, 16, 150, 5, 195, 15, 61, 23, 8033, 89, 3~
$ popasian
                       <int> 124, 9, 34, 1139, 6, 221, 0, 84, 6, 1596, 20, 7, ~
$ popother
                       <dbl> 96.71206, 66.38434, 96.57128, 95.25417, 90.19877,~
$ percwhite
                       <dbl> 2.57527614, 32.90043290, 2.86171703, 0.41225735, ~
$ percblack
                       <dbl> 0.14828264, 0.17880670, 0.23347342, 0.14932156, 0~
$ percamerindan
$ percasian
                       <dbl> 0.37675897, 0.45172219, 0.10673071, 0.48691813, 0~
                       <dbl> 0.18762294, 0.08469791, 0.22680275, 3.69733169, 0~
$ percother
                       <int> 43298, 6724, 9669, 19272, 3979, 23444, 3583, 1132~
$ popadults
                       <dbl> 75.10740, 59.72635, 69.33499, 75.47219, 68.86152,~
$ perchsd
$ percollege
                       <dbl> 19.63139, 11.24331, 17.03382, 17.27895, 14.47600,~
                       <dbl> 4.355859, 2.870315, 4.488572, 4.197800, 3.367680,~
$ percprof
$ poppovertyknown
                       <int> 63628, 10529, 14235, 30337, 4815, 35107, 5241, 16~
                       <dbl> 96.27478, 99.08714, 94.95697, 98.47757, 82.50514,~
$ percpovertyknown
$ percbelowpoverty
                       <dbl> 13.151443, 32.244278, 12.068844, 7.209019, 13.520~
$ percchildbelowpovert <dbl> 18.011717, 45.826514, 14.036061, 11.179536, 13.02~
                       <dbl> 11.009776, 27.385647, 10.852090, 5.536013, 11.143~
$ percadultpoverty
                       <dbl> 12.443812, 25.228976, 12.697410, 6.217047, 19.200~
$ percelderlypoverty
                       <int> 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0~
$ inmetro
$ category
                       <chr> "AAR", "LHR", "AAR", "ALU", "AAR", "AAR", "LAR", ~
```

1. Making a histogram to visualize the population density of counties

```
ggplot(midwest, aes(x = popdensity))+
  geom_histogram(binwidth = 10000, fill = "light blue")+
labs(title = "Population density in different counties", x = "Population density", y = "Fr
```





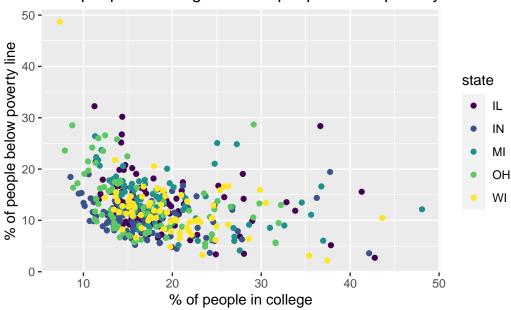
The distribution has a right skewed tail.

There seems to be a few counties with an extremely high population density at around 60000 and 90000. These are outliers, they are far away from most of the data and outside of the curve.

2. Create a scatterplot of the percentage of people with a college degree versus percentage below poverty

```
ggplot(midwest, aes(x=percollege, y=percbelowpoverty, color=state))+
   geom_point()+
   labs(title="% of people in college vs % of people below poverty line", x= "% of people
```

% of people in college vs % of people below poverty line

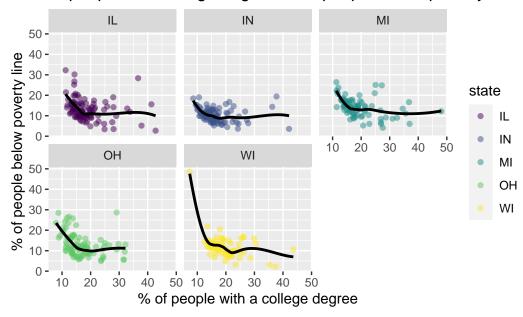


- 3. Describe what you observe in the plot from the previous exercise. In your description, include similarities and differences in the patterns across states.
- 4. Looking at the relationship between the number of poeple with a college degree and the number of poeple below the poverty line by state.

```
#plotting data on a scatter plot, diving grids based on the state
ggplot(midwest, aes(x=percollege, y=percbelowpoverty, color=state))+
  geom_point(alpha = .5)+
  facet_wrap(~state) +
  geom_smooth(se = FALSE, color = "black")+
  labs(title="% people with college degree vs % people below poverty line", x= "% of
```

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

% people with college degree vs % people below poverty line

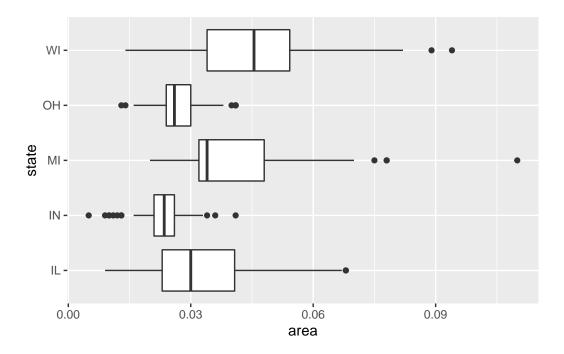


#color black for the lines makes the line more visible

Which plot do you prefer - this plot or the plot in Ex 2? Briefly explain your choice.

5. Looking at the difference in area between states.

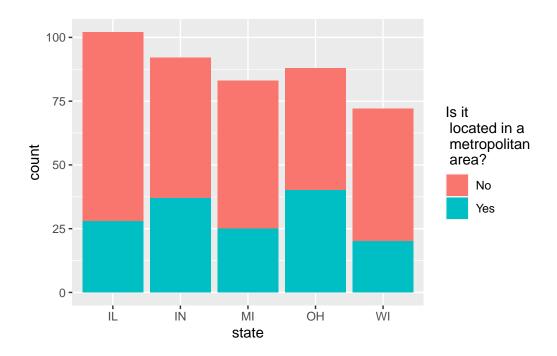
```
#plotting the data on side by side boxplots
ggplot(midwest, aes(x=area, y=state))+
geom_boxplot()
```



- Describe what you observe from the plot.
- Which state has the single largest county? How do you know based on the plot?
- 6. Exploring whether most counties in the chosen states are located in a metropolitan area or not

```
#data wrangling code
midwest <- midwest |>
   mutate(metro = if_else(inmetro == 1, "Yes", "No"))

#creating a segmented bar chart for the data
ggplot(midwest, aes(x=state, fill= metro))+
   geom_bar()+
   labs(fill= "Is it \n located in a \n metropolitan \n area?")
```



7. Looking at whether people with a college degree tend to live in denser areas

```
#reproducing the scatter diagram
    ggplot(midwest, aes(x=percollege, y=popdensity, color=percbelowpoverty))+
        geom_point(size=2, alpha=0.5)+
        facet_wrap(~state) +
        labs(title="Do people with college degrees tend to live in denser areas?", x= "% col
```

Do people with college degrees tend to live in denser areas?

