Homework 4

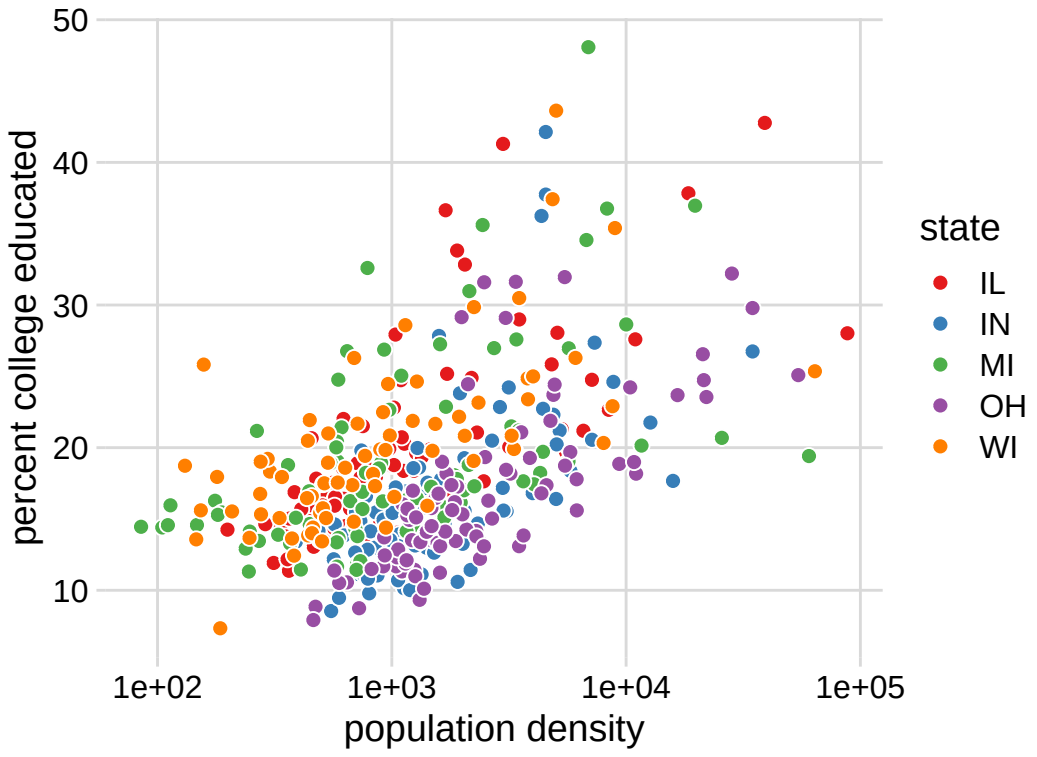
Wenxuan Zhou wz4388

This homework is due on Feb. 15, 2021 at 11:00pm. Please submit as a pdf file on Canvas.

Problem 1: (5 pts) We will work with the `midwest` dataset provided by **ggplot2**. See here for details: https://ggplot2.tidyverse.org/reference/midwest.html

For the following plot, add an appropriate color scale from the **colorspace** package and explicitly set the palette you want to use. Also add an appropriate theme that works well with the color scale you have chosen. Explain in 2-3 sentences your choice of color scale and theme.

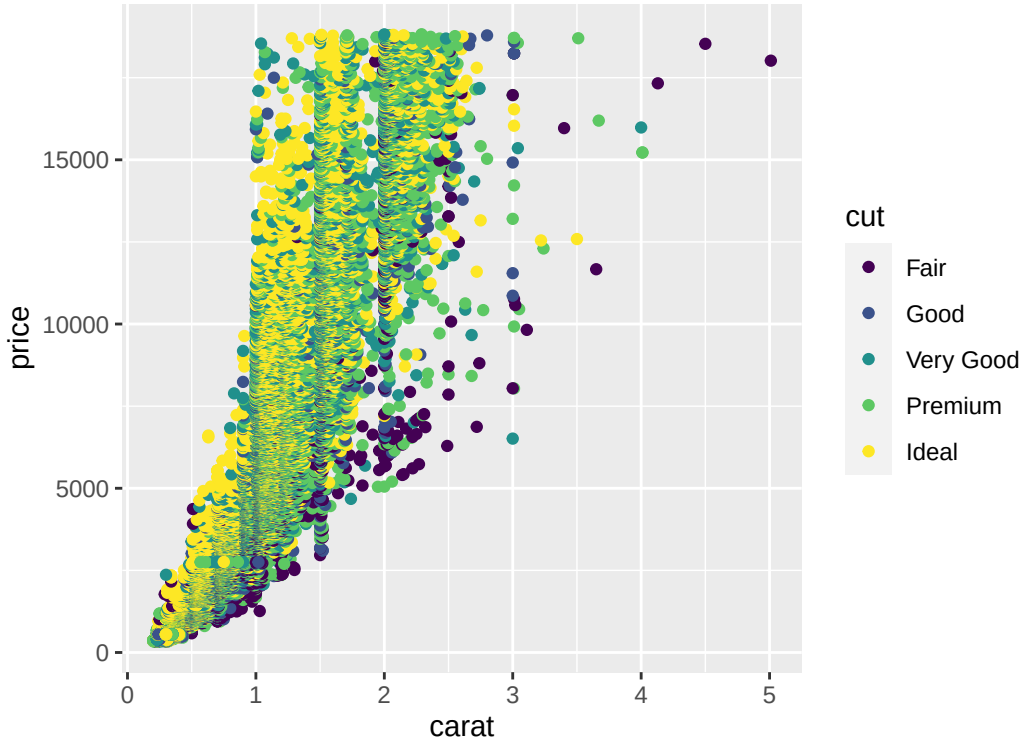
```
ggplot(midwest, aes(popdensity, percollege, fill = state)) +
  geom_point(shape = 21, size = 2.5, color = "white", stroke = 0.5) +
  scale_x_log10(name = "population density") +
  scale_y_continuous(name = "percent college educated") +
  scale_fill_brewer(palette = 'Set1') +
  theme_minimal_grid()
```



I choose to use scale_fill_brewer() because its aesthetic is fill and has qualitative palette type. I use palette Set1 because the color for each state is distinct from each other, which makes the data points clear to read. theme_minimal_grid() is used because it makes the plot simple and easy to read.

Problem 2: (2 pts) The following is a plot of the `diamonds` dataset provided by **ggplot2**. See here for details: https://ggplot2.tidyverse.org/reference/diamonds.html

```
ggplot(diamonds, aes(carat, price, color = cut)) +
  geom_point()
```

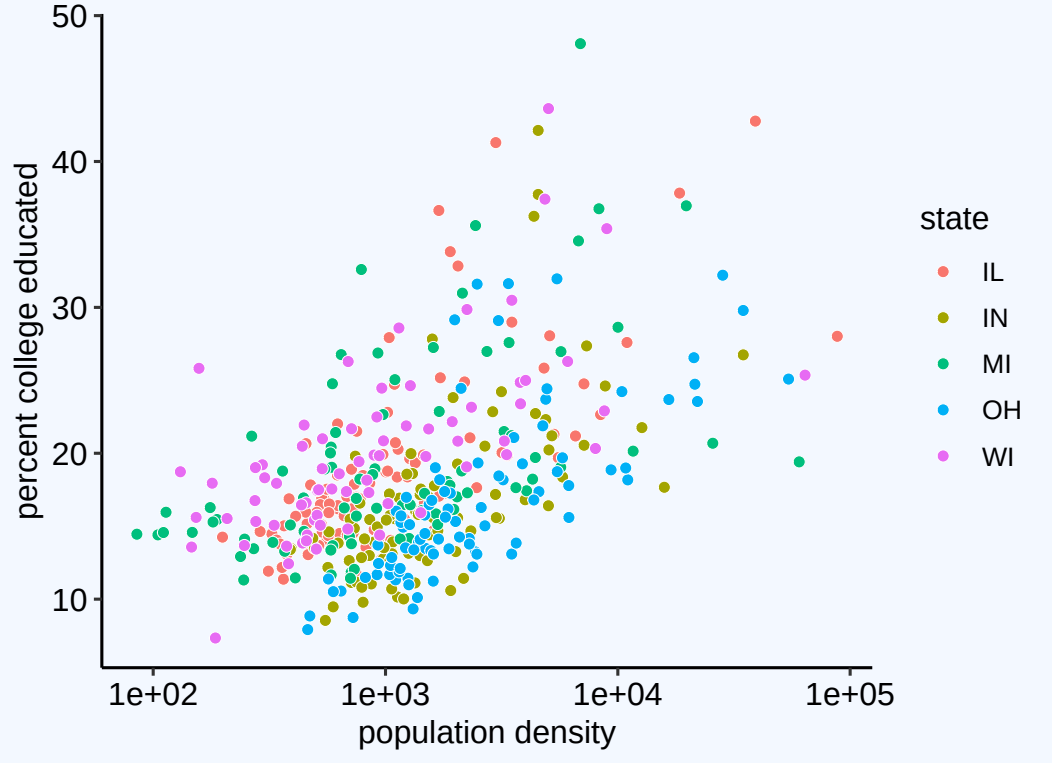


As you can see, if we don't specify a color scale ggplot chooses the viridis scale by default. Explain in 2-3 sentences why the viridis scale is an appropriate choice here.

The viridis scale is an appropriate choice here because it has a sequential palette type from blue to yellow. The quality of the cut can also be considered sequential (from fair to ideal). Even though the color scale is sequential, data points for each cut quality can also be easily distinguished, and ideal cuts pop out in the plot.

Problem 3: (3 pts) For this problem, we will go back to the `midwest` dataset. In the following plot, you may notice that the axis tick labels are smaller than the axis titles, and also in a different color (gray instead of black). Make the axis tick labels the same size (`size = 12`) and give them the color black (`color = "black"`). Then, set the entire plot background to the color `"#F3F8FF"`. Make sure there are no white areas remaining, such as behind the plot panel or under the legend.

```
ggplot(midwest, aes(popdensity, percollege, fill = state)) +
  geom_point(shape = 21, size = 2, color = "white", stroke = 0.2) +
  scale_x_log10(name = "population density") +
  scale_y_continuous(name = "percent college educated") +
  theme_classic(12) +
  theme(
    axis.text = element_text(size = 12, color = 'black'),
    plot.background = element_rect(fill = "#F3F8FF"),
    panel.background = element_rect(fill = "#F3F8FF"),
    legend.background = element_rect(fill = "#F3F8FF")
  )
```


50
40
30
20
10
percent college educated
1e+02
1e+03
1e+04
1e+05
population density
state
IL
IN
MI
OH
WI