MPA 634  
Data Science and R for Administrators  
Homework #2

Aesthetic Mappings and Facets

1. Problems 1 – 6 from Exercises 3.3.1
2. Problem 4 from Exercises 3.5.1

3.3.1 – 1) What’s gone wrong with this code? Why are the points not blue?

The points are not all blue because the way to set that aesthetic manually is to set the aesthetic by name as an argument of the geom function, not the aes function. The code should be:  
geom\_point(mapping = aes(x=displ, y=hwy), color = “blue”)

3.3.1 – 2) Which variables in mpg are categorical? Which variables are continuous? (Hint: type ?mpg to read the documentation for the dataset). How can you see this information when you run mpg?

Categorical: Manufacturer, Model, Year, Cylinder, Transmission, Drive, Fuel Type, Class

Continuous: Displacement (engine displacement could theoretically be any volume), City MPG, Hwy MPG

You can see this information easily by simply clicking on the mpg in the Data window to produce a table that can be easily interpreted as a tab in the source code terminal.

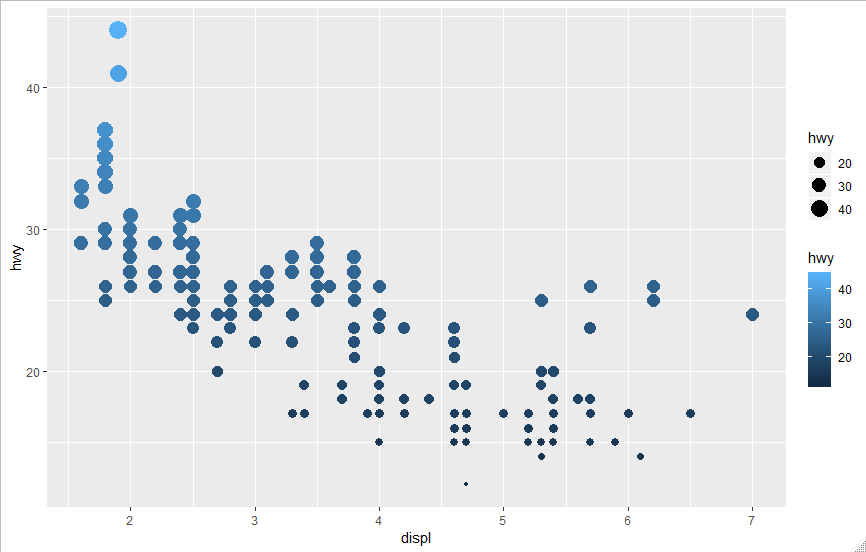
3.3.1 – 3) Map a continuous variable to color, size, and shape. How do these aesthetics behave differently for categorical vs. continuous variables?

A continuous variable cannot be mapped to a shape, but it can be mapped to color and size. Here’s the code and the resulting map.

#Map a continuous variable (hwy mpg) to color, size, and shape.

ggplot(data = mpg) +

geom\_point(mapping = aes(x=displ, y=hwy, color = hwy, size = hwy))#, shape = hwy))

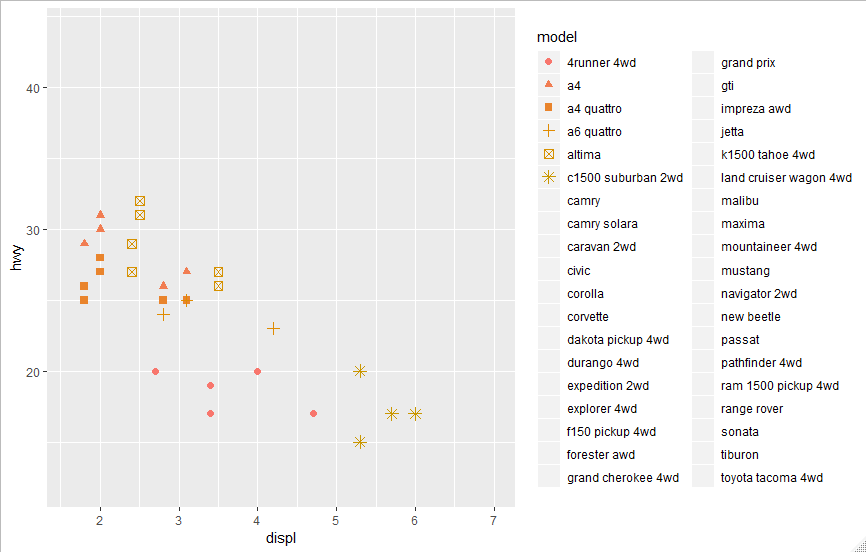


For categorical variables, I used Model and got the following plot. It can plot discrete values for color, size, and shape, but it’s more limited in the number of shapes and in this case had to remove data.

#Map a categorical variable (displ) to color, size, and shape.

ggplot(data = mpg) +

geom\_point(mapping = aes(x=displ, y=hwy, color = model, size = model, shape = model))



Basically, because there are theoretically infinite numbers of possibilities for continuous variables, you cannot map them with a discrete value such as shape. Color and size work because they too can be continuous. In the case of categorical variables, they can be mapped somewhat easily using both discrete and continuous properties, although it is not advised and is limited by the number of discrete possibilities in the program, as seen in the number of shape options not being the same as the number of different models that needed to be represented.

3.3.1 – 4) What happens if you map the same variable to multiple aesthetics?

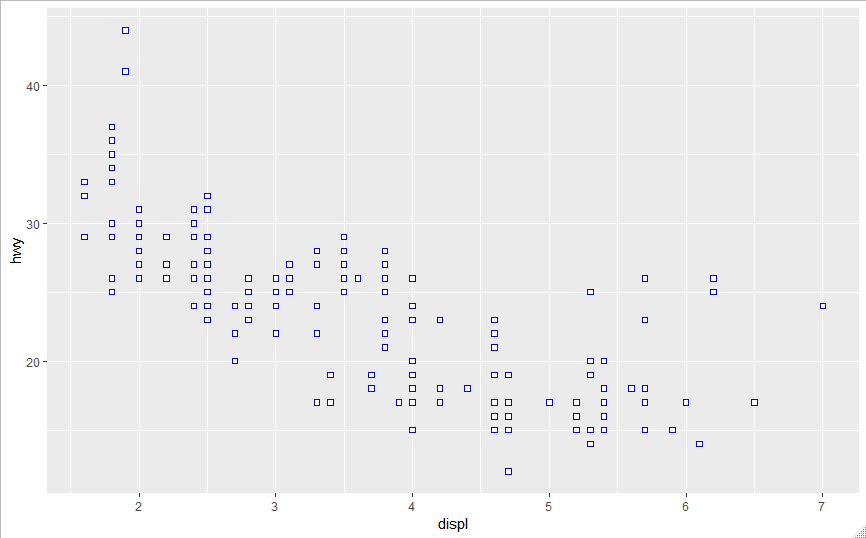
Depending on the type of variable, it may not plot. However, for the most part, it should plot fine but have multiple aesthetics showing the same thing. For instance, in the previous question, hwy was mapped using size and color. So you can see the difference in hwy mpg for each car two different ways, which may be confusing and slightly unnecessary.

3.3.1 – 5) What does the stroke aesthetic do? What shapes does it work with? (Hint: use ?geom\_point)

Stroke can be used to modify the width of a border for shapes with borders (shapes 21-24)

ggplot(data = mpg) +

geom\_point(mapping = aes(x=displ, y=hwy), shape = 22, color = "blue", stroke = 1, fill = "White")



3.3.1 – 6) What happens if you map an aesthetic to something other than a variable name, like aes(colour = displ < 5)? Note, you’ll also need to specify x and y.



As seen in the plot above, mapping an aesthetic to something other than a variable, like an expression, still plots but uses the aesthetic to display the results of the expression. In this case, the expression displ < 5 was mapped by having blue represent the points where that expression is true and red represent where that expression is false. Essentially, it will apply the aesthetic to those points where the expression is true and differentiate those where it is not true.

3.5.1 – 4) What are the advantages to using faceting instead of the colour aesthetic? What are the disadvantages? How might the balance change if you had a larger dataset?

Faceting has a couple advantages. First, it allows you to see each set of data separately without having to worry about overlapping. Second, it can give you a better idea of trends for each variable. Third, if the range of values on some sets of data are vastly different from others, breaking them apart using faceting will allow a better view of the data than if they are all on the exact same plot broken apart using the color.

Some disadvantages of faceting include the limited capability to handle a large number of different variables. If you want to look at more than a few variables, then faceting can be difficult as you’d have to look at a ton of different plots separately. Second, faceting makes a direct comparison of different variables difficult as the data is not overlaid on each other. Finally, faceting can sometimes just be somewhat difficult to read with all of the different scales and tons of plots.

With a larger dataset, faceting becomes more difficult to use as it produces more and more plots if there are more variables. Also, the ranges of the data may become large and all over the place. In those situations, using color could make it easier to interpret the data in one go.

