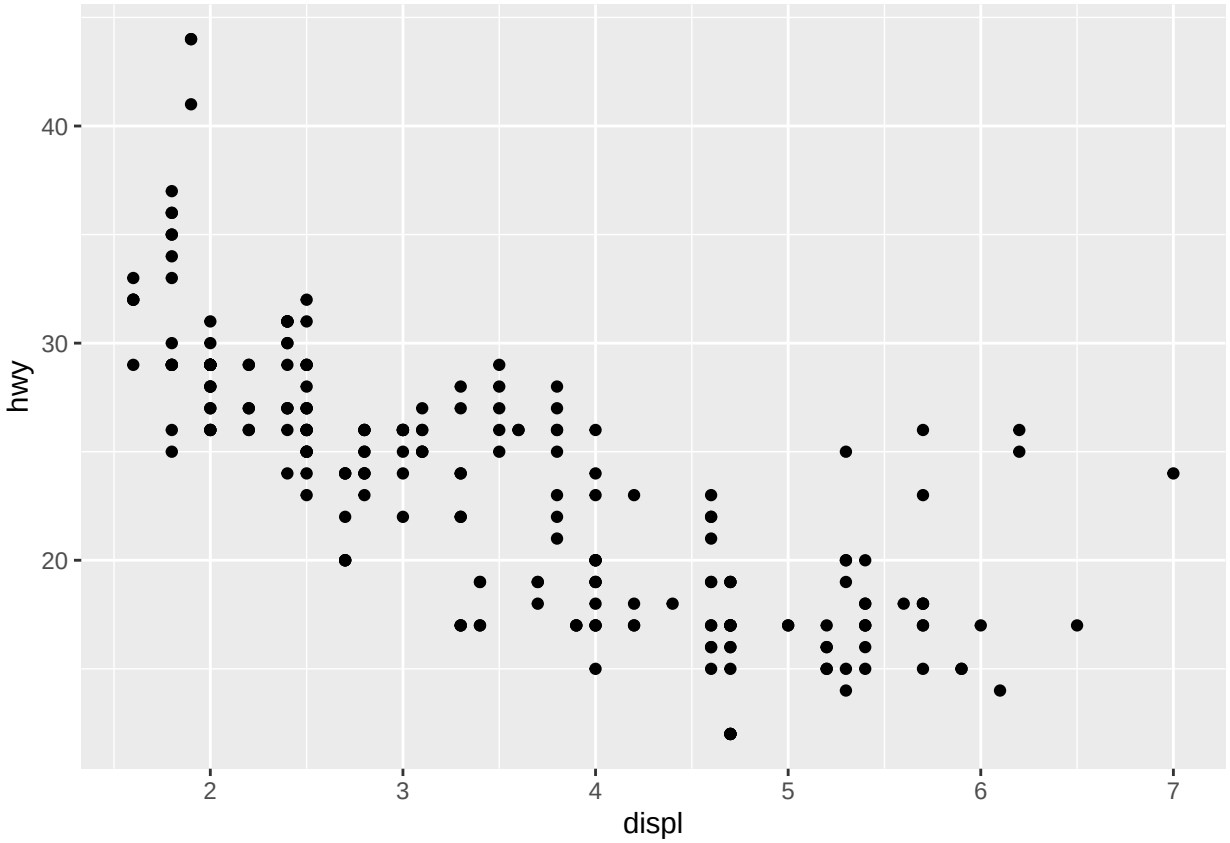
Chapter 3

Icaro Franco Hernandes

2022/09/23

```
library(ggplot2)
ggplot2::ggplot(data=ggplot2::mpg)+
  geom_point(aes(x=displ, y=hwy))
```



Placing the *aes* inside the ggplot we do not need to restate the variables in the geometry parts. All the aesthetic changes will be made considering the declared dataset. If instead you place the *aes* in the geometry segment, you'll have to change accordly to what type of graph you want.

Exercises 3.2.4

Exercise 1

Run $ggplot(data = mpg)$. What do you see?

```
ggplot2::ggplot(data=mpg)
```


We see nothing. Adding just the *ggplot* creates a canvas to which we could add layers upon with further codes, like *geom_point*. —

Exercise 2

How many rows are in mpg? How many columns?

```
nrow(mpg)
```

```
## [1] 234
```

```
ncol(mpg)
```

```
## [1] 11
```

Exercise 3

What does the *drv* variable describe? Read the help for *?mpg* to find out.

```
?mpg
```

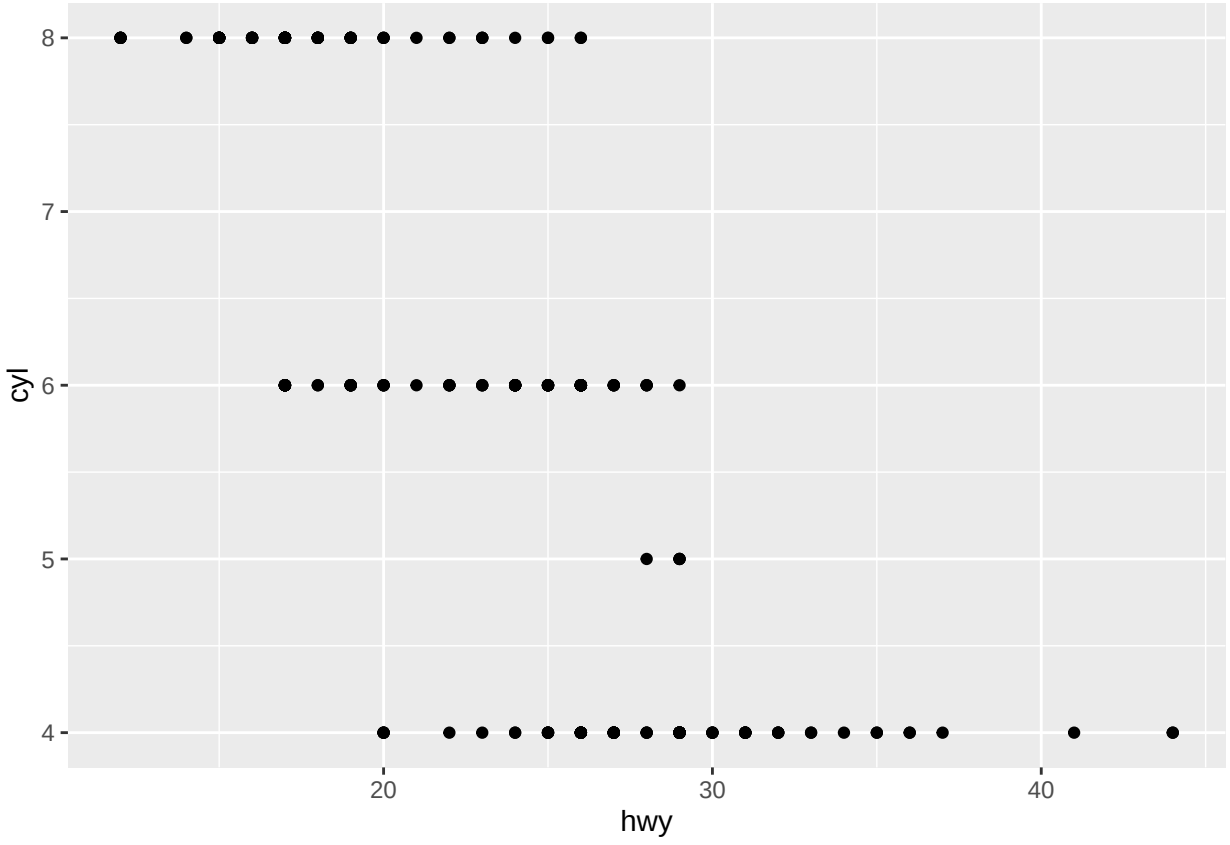
```
## starting httpd help server ... done
```

Adding the question mark before any command takes us to the Help tab. According to the documentation of *mpg* the *drv* variable gives us what type of drive train the car uses, wether is front, rear ow four-wheel drive.

Exercise 4

Make a scatterplot of *hwy* vs *cyl*.

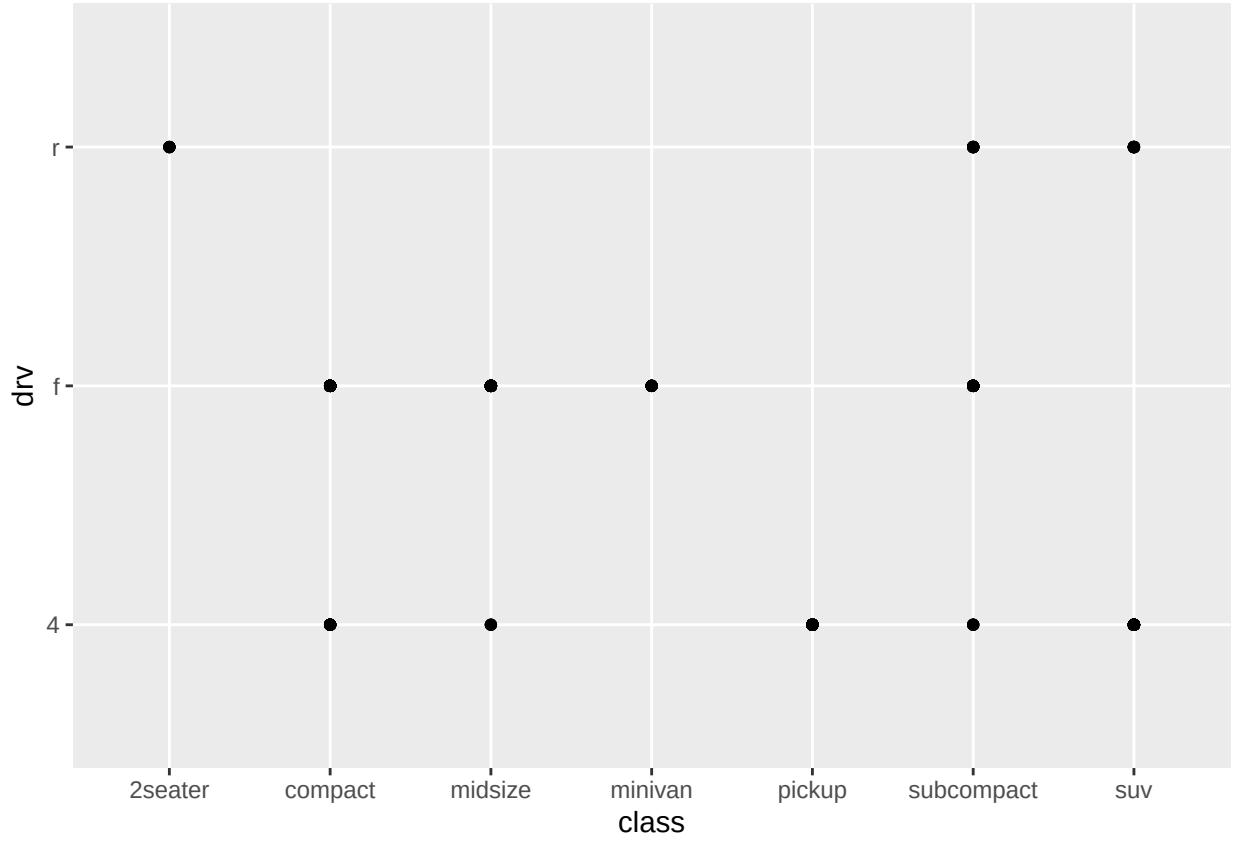
```
ggplot2::ggplot(data=mpg, aes(x=hwy, y=cyl))+
  geom_point()
```



Exercise 5

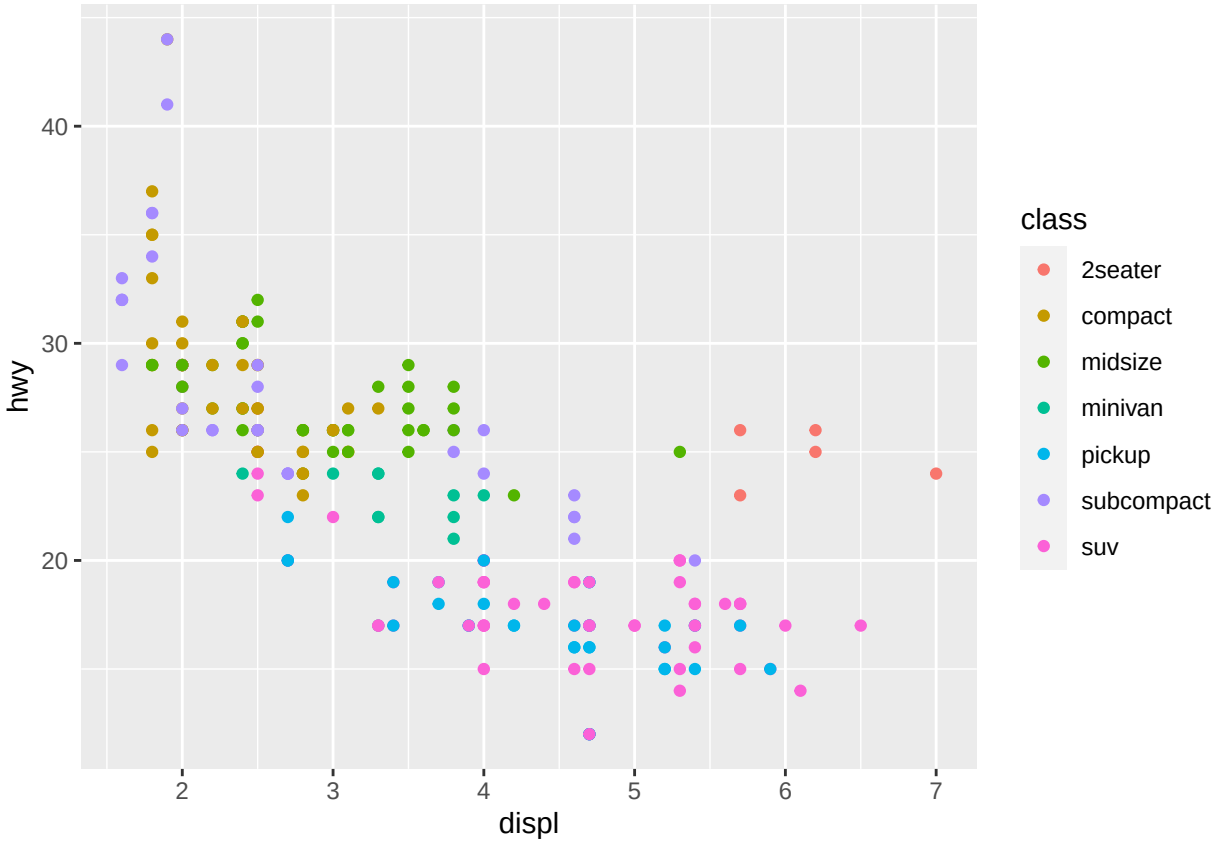
What happens if you make a scatterplot of $class$ vs drv? Why is the plot not useful?

```
ggplot2::ggplot(data=mpg,aes(x=class, y=drv))+
  geom_point()
```



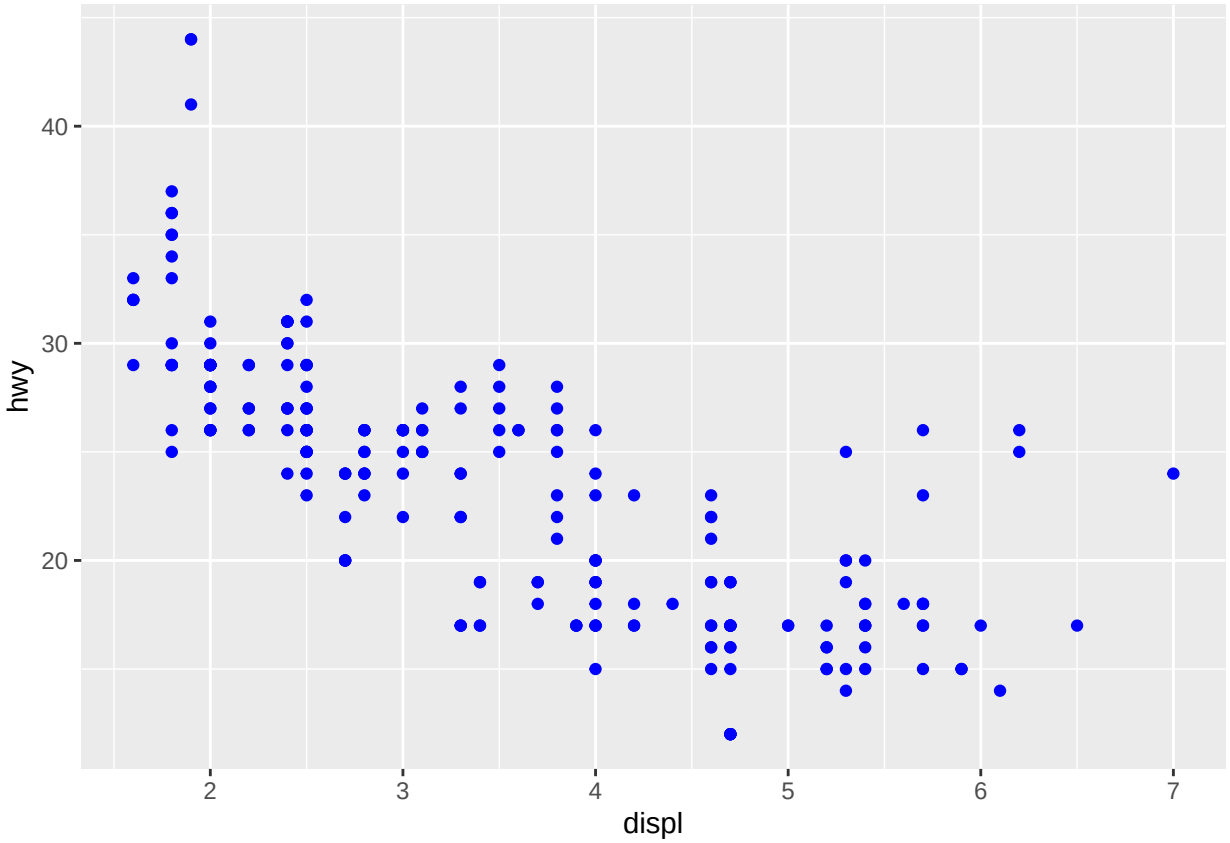
This scatter plot is not useful since it does not indicate any pattern. No inference can be made with this graph.

```
ggplot2::ggplot(data=mpg)+
  geom_point(aes(x=displ,y=hwy, color=class))
```



It is not a good practice to map an unordered variable to a ordered aesthetic (like in the example $size = class$). $alpha = variable$ is for transparency of the points and $shape = variable$ is for each variable an observation has a different shape in the graph.

```
ggplot2::ggplot(data=mpg)+
  geom_point(aes(x=displ,y=hwy), color="blue")
```



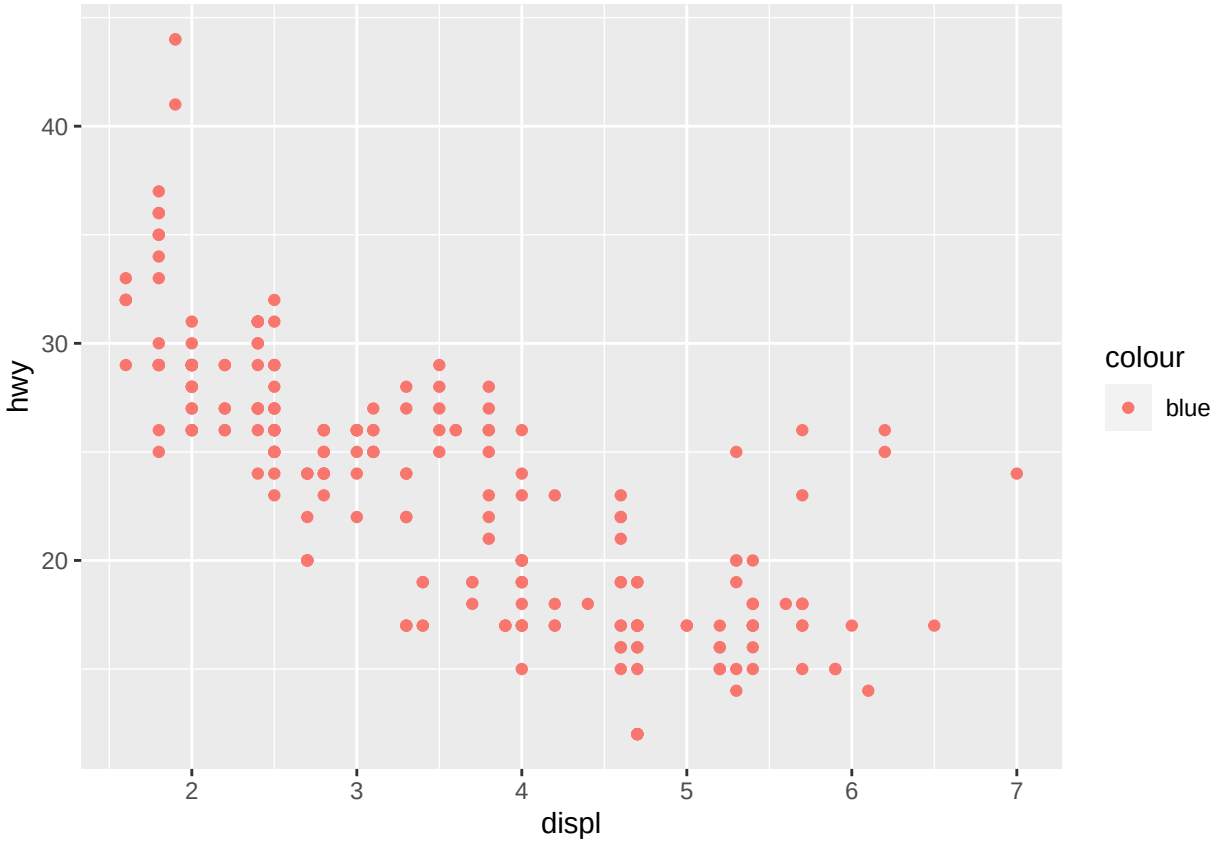
This way, all the points will be blue. Since the command is outside the *aes*() "the color does not convey information about a variable".

Exercises 3.3.1

Exercise 1

What's gone wrong with this code? Why are the points not blue?

```
ggplot(data = mpg) +
  geom_point(aes(x = displ, y = hwy, color = "blue"))
```



The points are not blue since the color command is inside the *aes*() function. That would be corret if we wanted to segment the data by some variable. If we just want to paint all the points blue we'd just have to put the command outside *aes*()

Exercise 2

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

```
head(mpg)
```

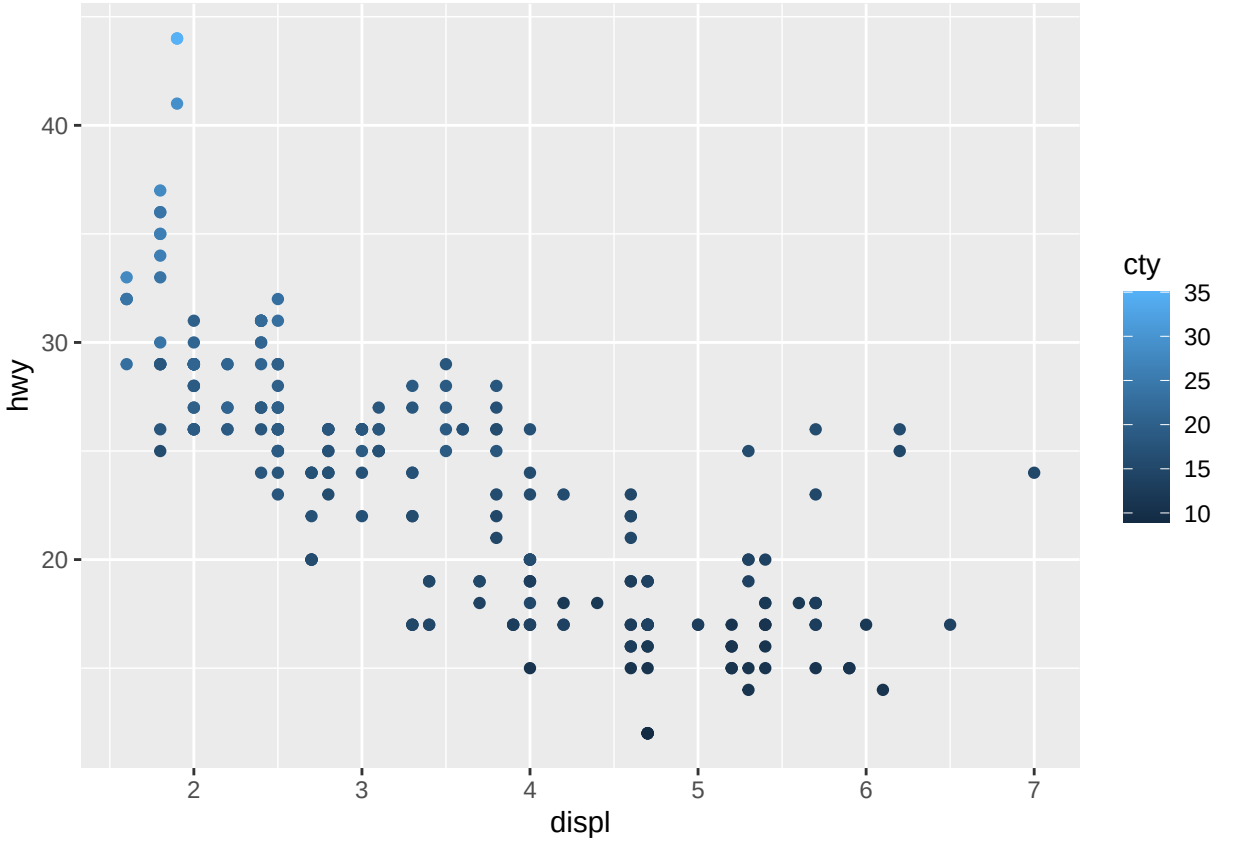
```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

The only continuous variables are *cty* and *hwy*. These variables regard the fuel consumption in the city and on the highway.

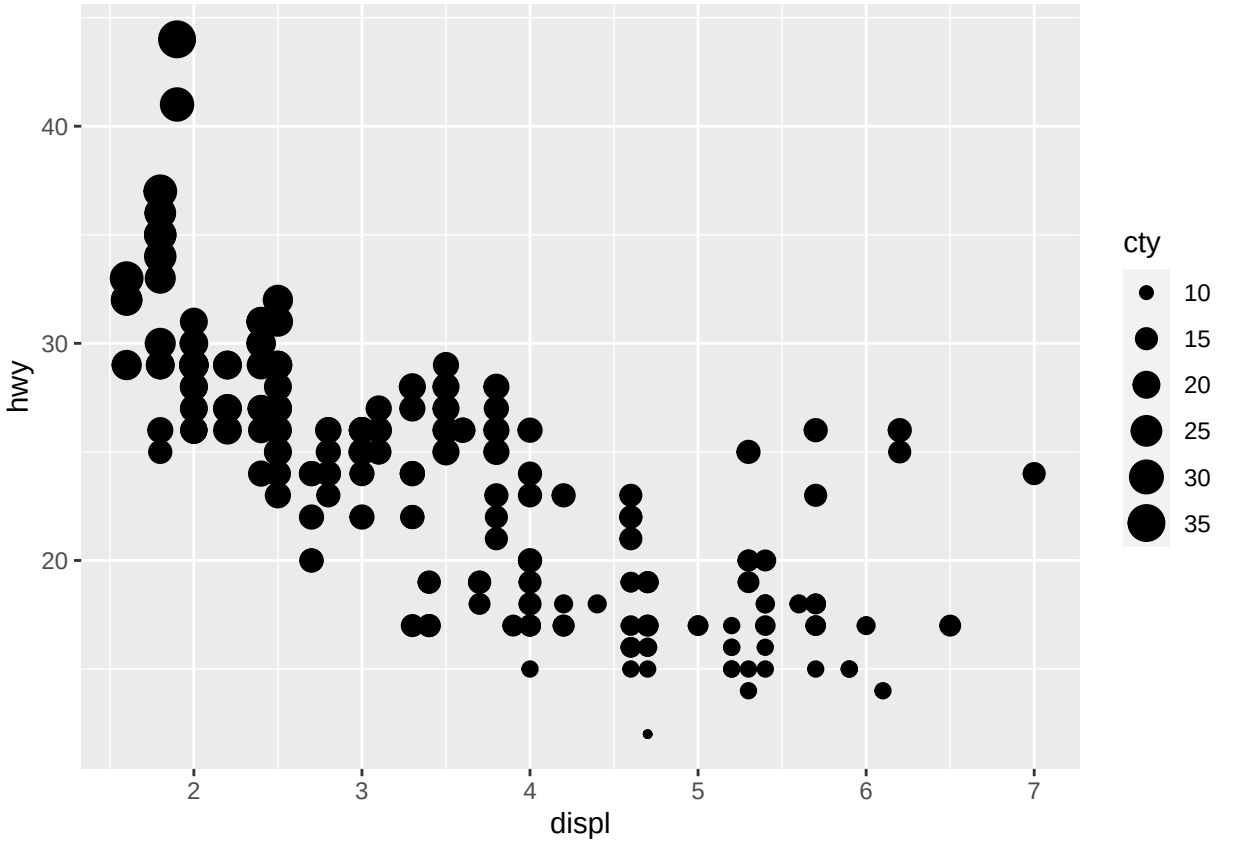
Exercise 3

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

```
ggplot2::ggplot(data=mpg)+
  geom_point(aes(x=displ, y=hwy, color=cty))
```




```
ggplot2::ggplot(data=mpg)+
  geom_point(aes(x=displ, y=hwy, size=cty))
```

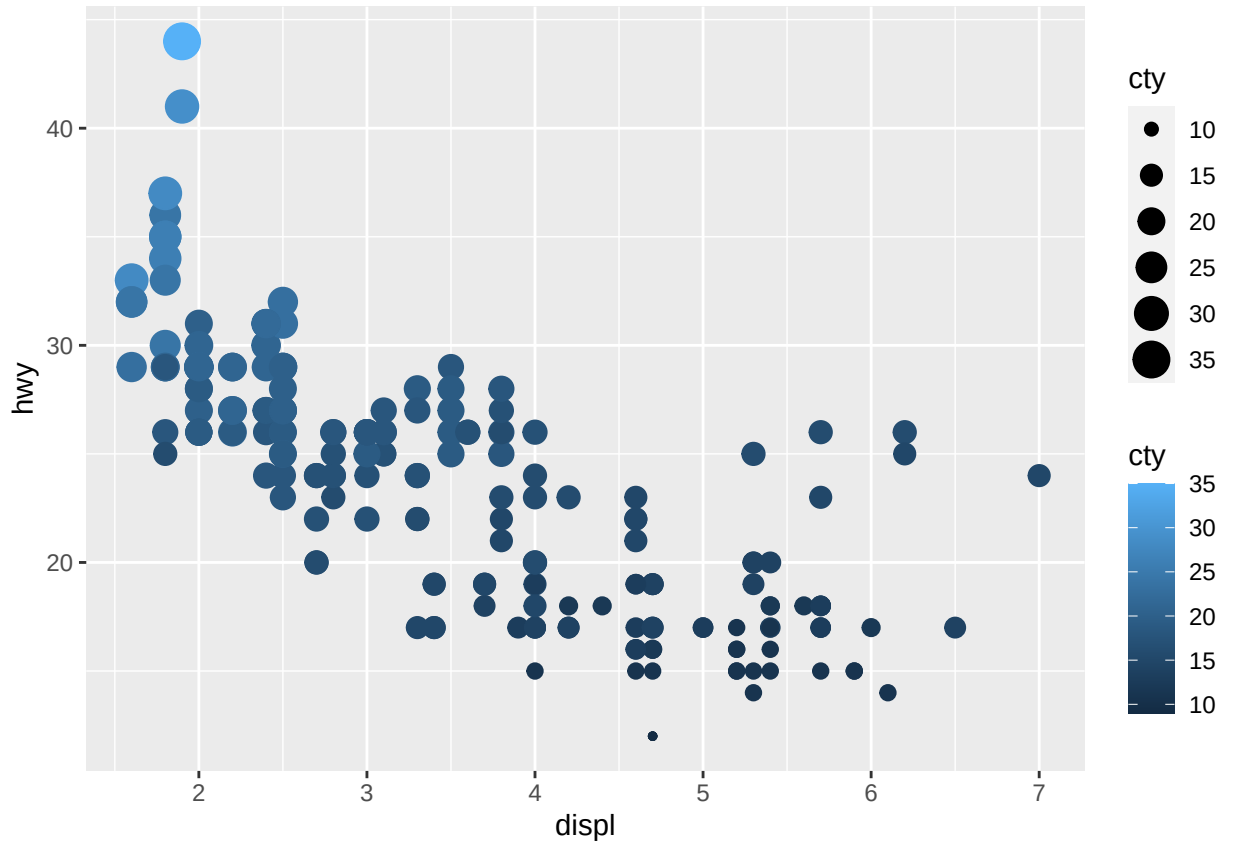


The last graph proposed cannot be plotted since "a continuous variable cannot be mapped to shape". The way they behave differently in the way that a continuous variable will give a gradient for colors and sizes while a discrete variable will have a different color for each class.

Exercise 4

What happens if you map the same variable to multiple aesthetics?

```
ggplot2::ggplot(data=mpg)+
  geom_point(aes(x=displ, y=hwy, color=cty, size=cty))
```

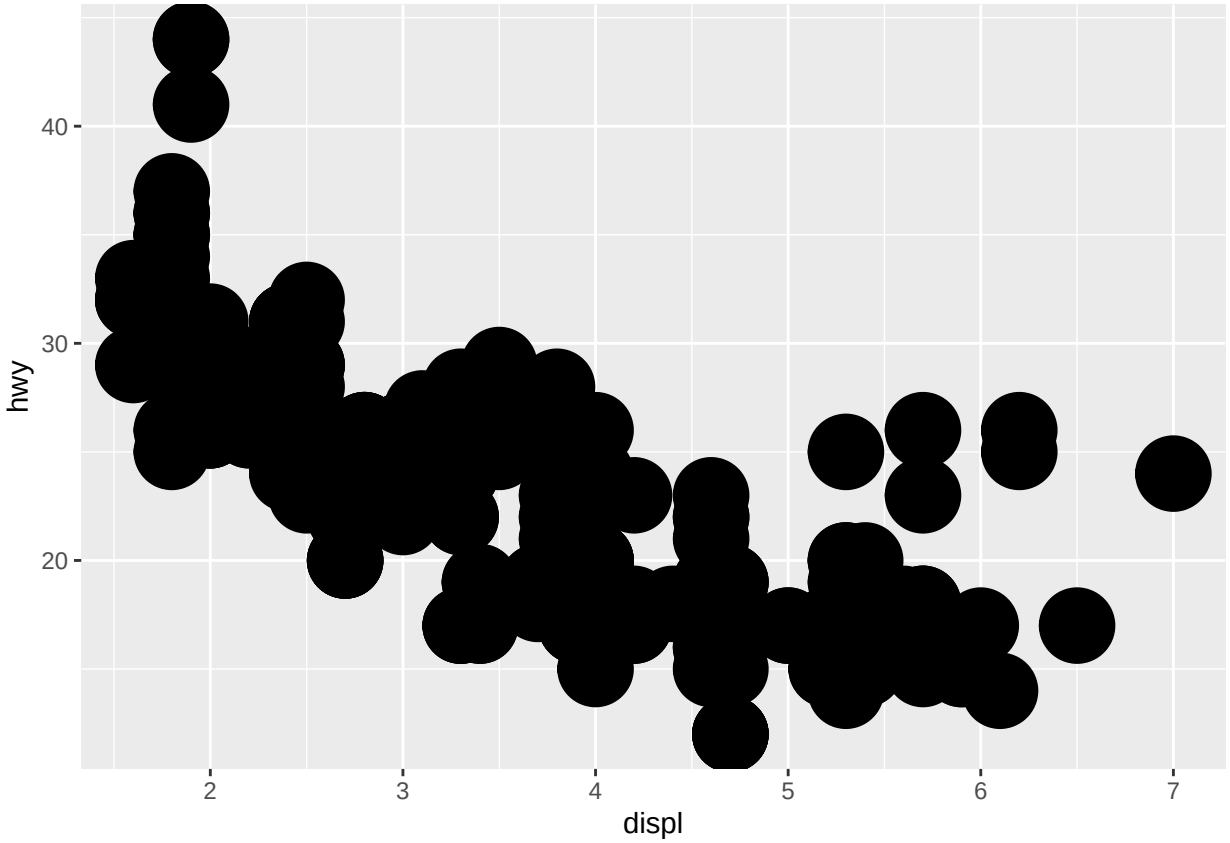


They get combined (is this the answer they were expecting?)

Exercise 5

What does the *stroke* aesthetic do? What shapes does it work with? (Hint: use *?geom_point*)

```
ggplot2::ggplot(data=mpg)+
  geom_point(aes(x=displ, y=hwy, stroke=9))
```

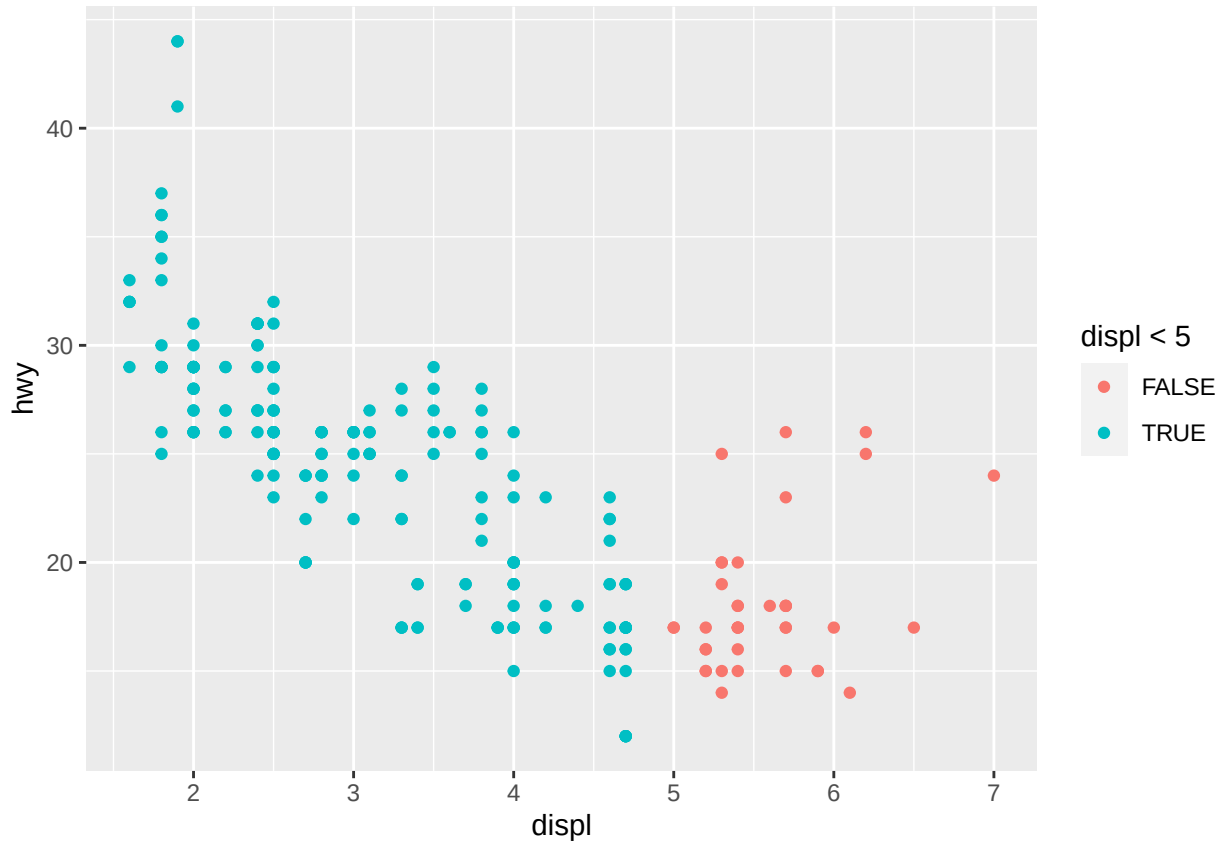


The stroke changes the width of the border for each observation.

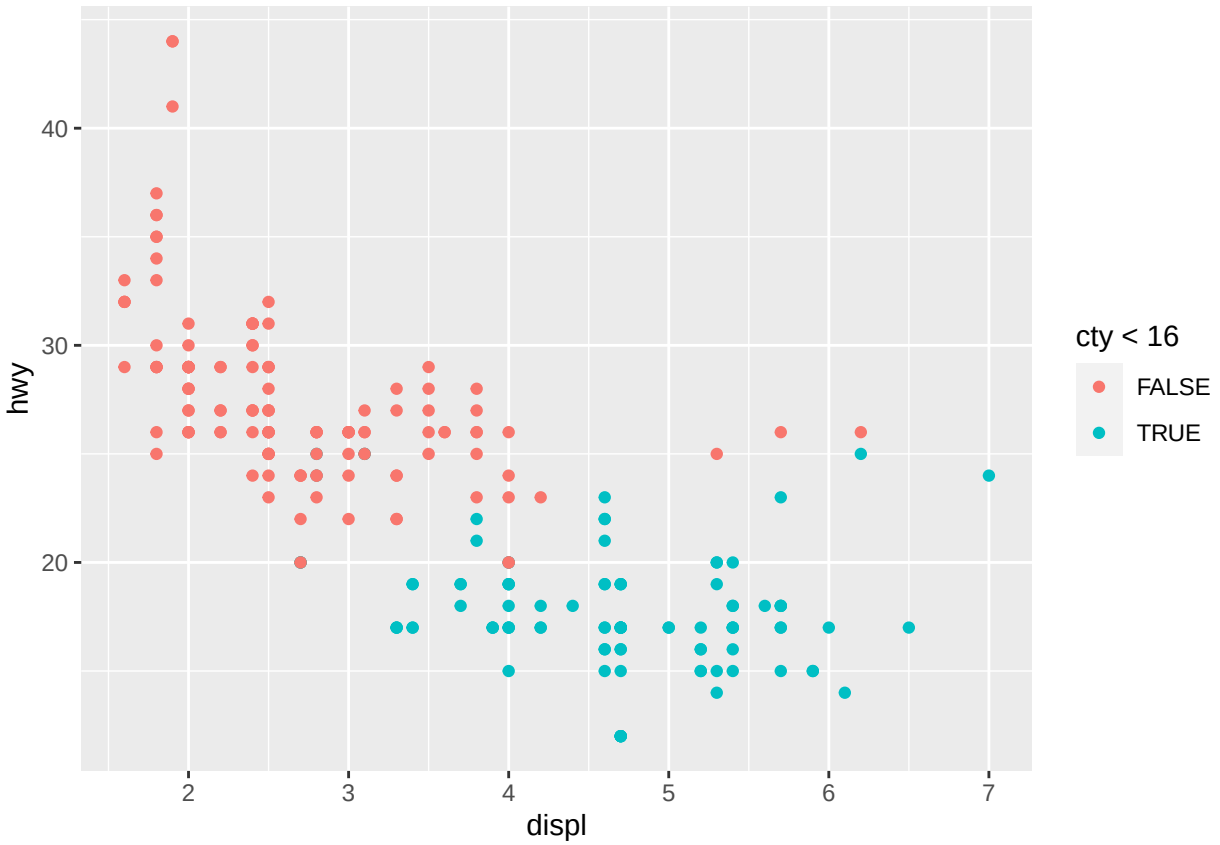
Exercise 6

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

```
ggplot2::ggplot(data=mpg)+
  geom_point(aes(x=displ,y=hwy, color=displ<5))
```




```
ggplot2::ggplot(data=mpg)+
  geom_point(aes(x=displ,y=hwy, color=cty<16))
```



This way we will "create" two classes to be colored. Those cars with less than 5 cylinders will have a color, and the cars with more than 5 cylinders will have another.