

ST537 - Lab1

Nora Quick

3.1

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5    v purrr  0.3.4
## v tibble  3.1.6    v dplyr  1.0.7
## v tidyr   1.1.4    v stringr 1.4.0
## v readr   2.0.2    v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

No error. I have it installed already. -NOTE: install tidyverse if needed

“ggplot2::ggplot()” specifies ggplot withing ggplot2

3.2

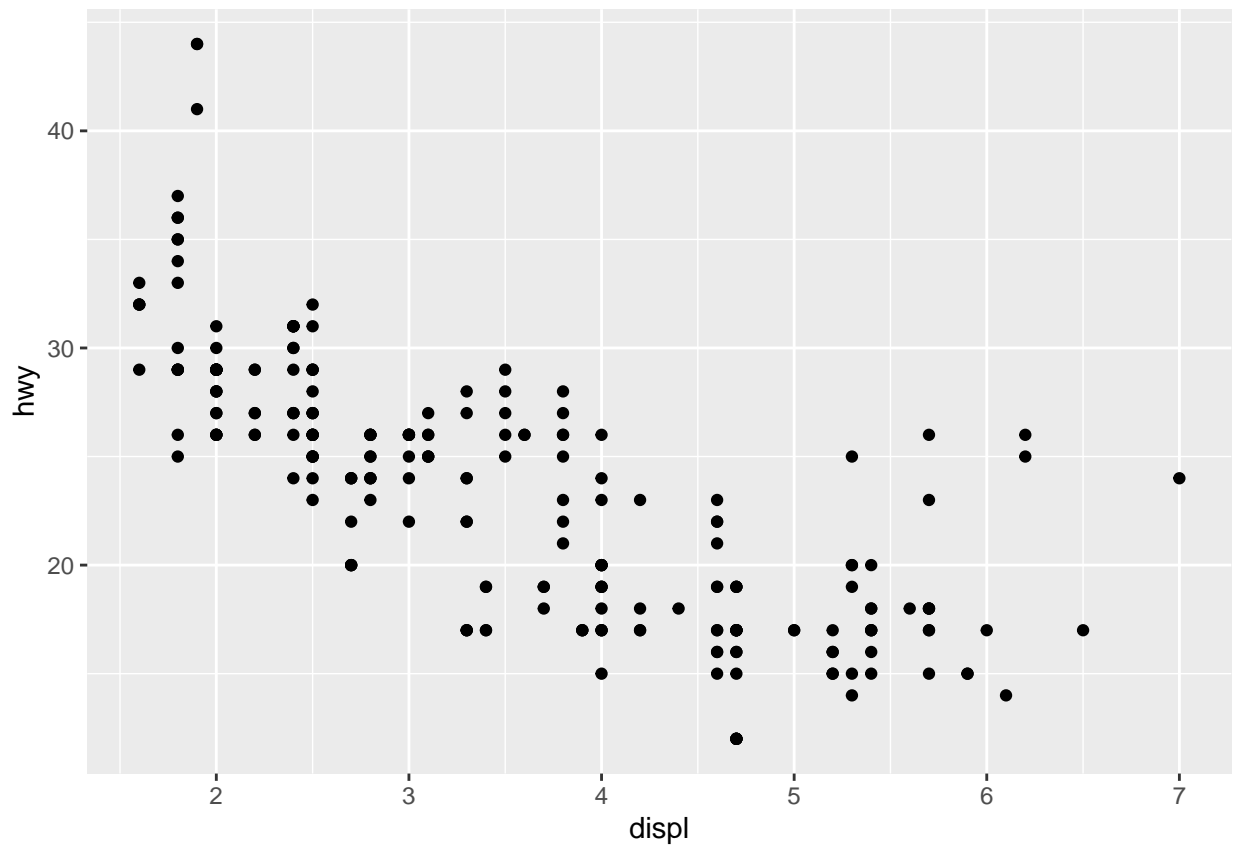
```
#mpg
```

```
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~
```

```
?mpg
```

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy))
```



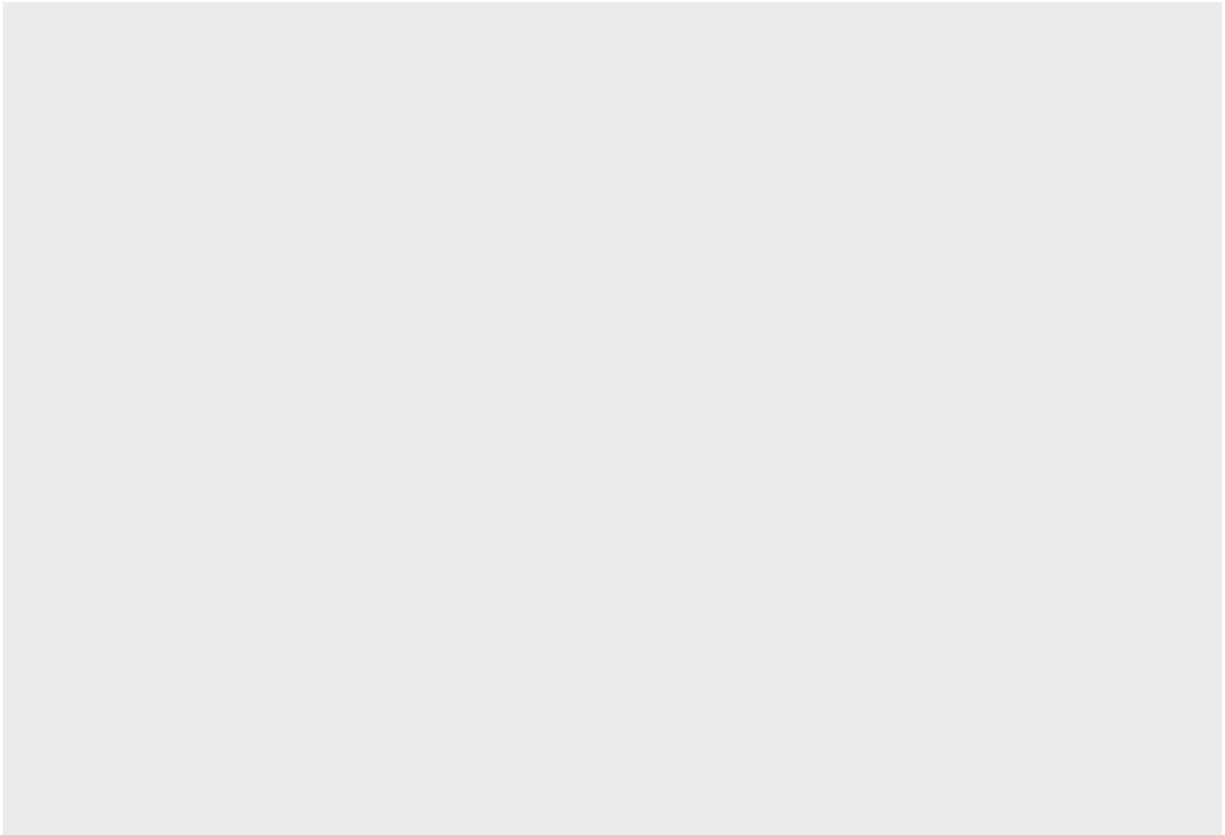
`geom_point()` == scatterplot

REUSABLE TEMPLATE: `ggplot(data =) + (mapping = aes())`

– brackets with dataset and geom function or collection of mappings

##3.2.4 ### 1.Run `ggplot(data = mpg)`. What do you see?

```
ggplot(data = mpg)
```



I see a blank graph as described in an earlier section.

2. How many rows are in mpg? How many columns?

```
?mpg
```

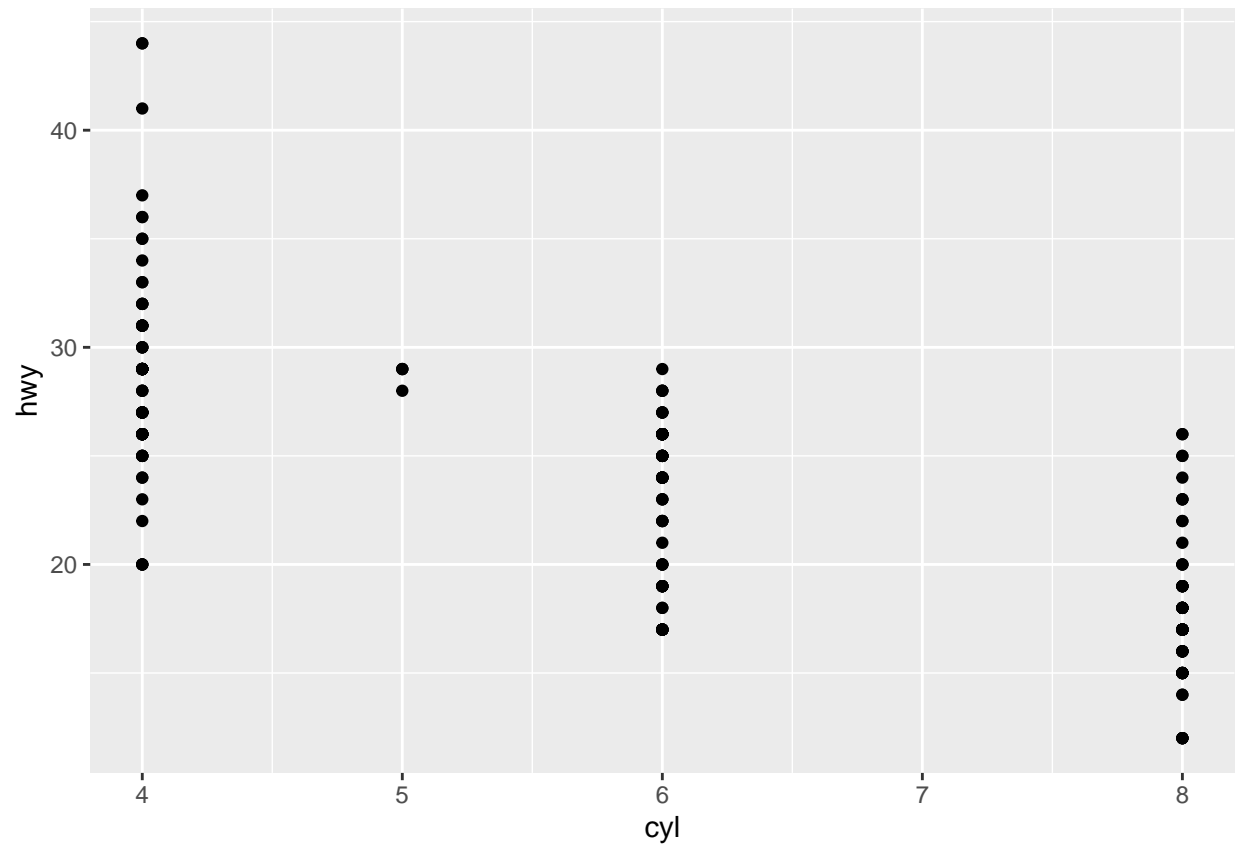
Help says there are 234 rows and 11 columns.

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

“drv” is the type of drive train where f is the front-wheel drive, r is the rear wheel drive, and 4 is four wheel drive.

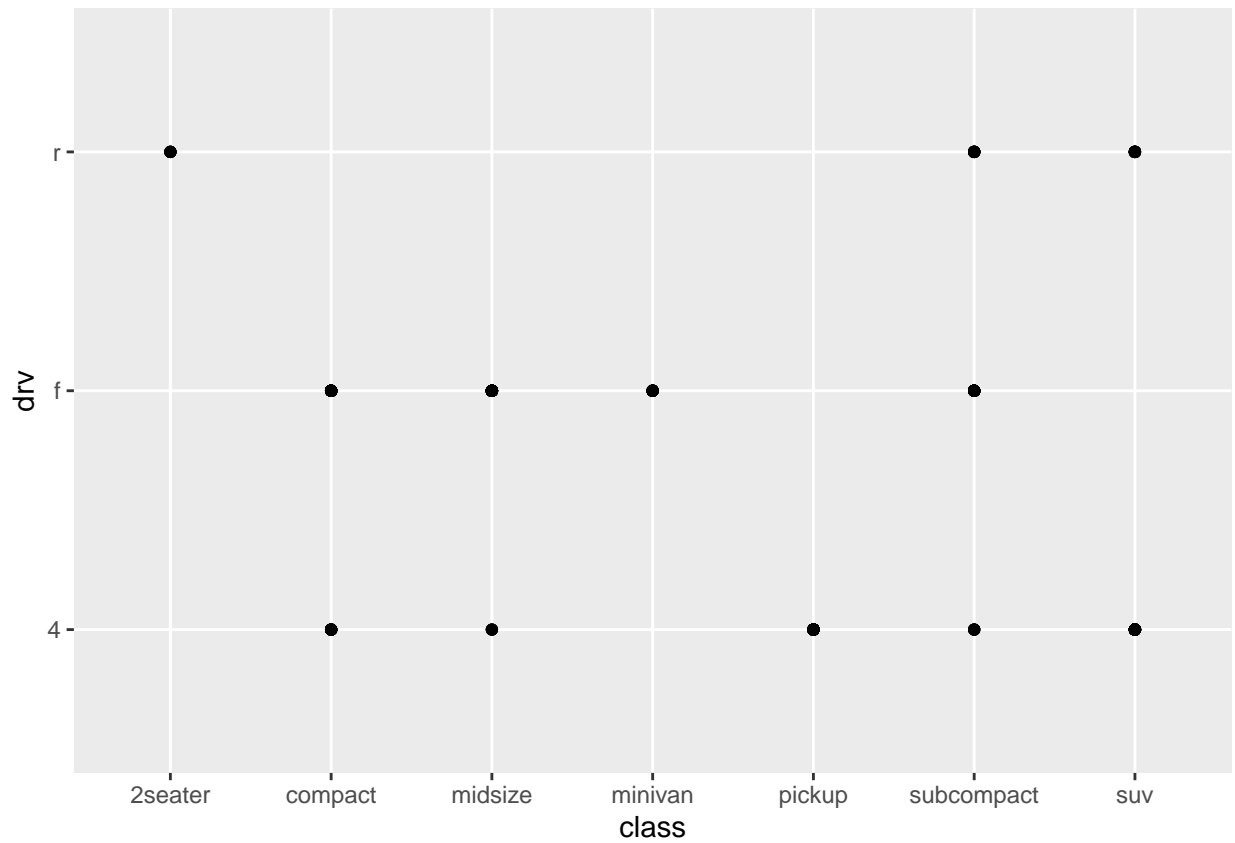
4. Make a scatterplot of hwy vs cyl.

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



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

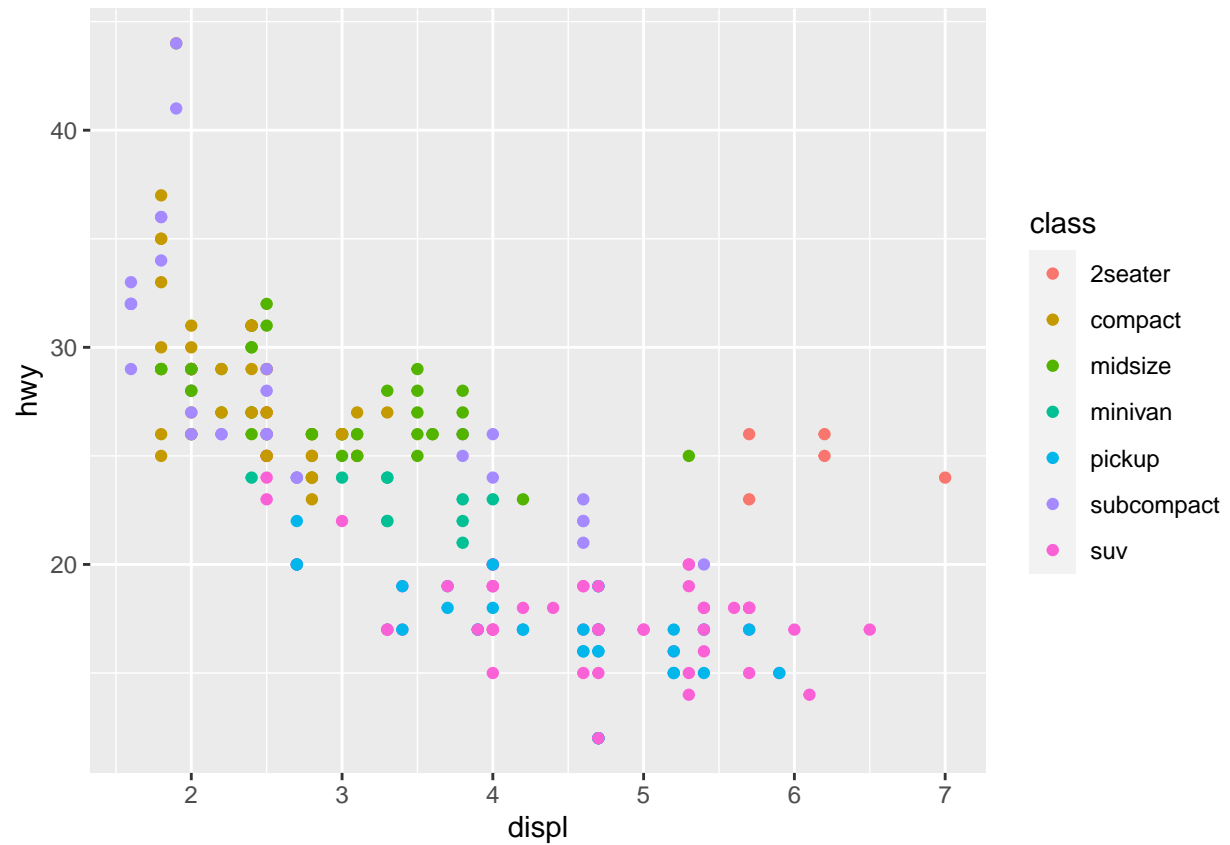
```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = class, y = drv))
```



The above is what happens when a scatterplot of class vs drv is made. This is not helpful because it is two descriptive/categorical variables.

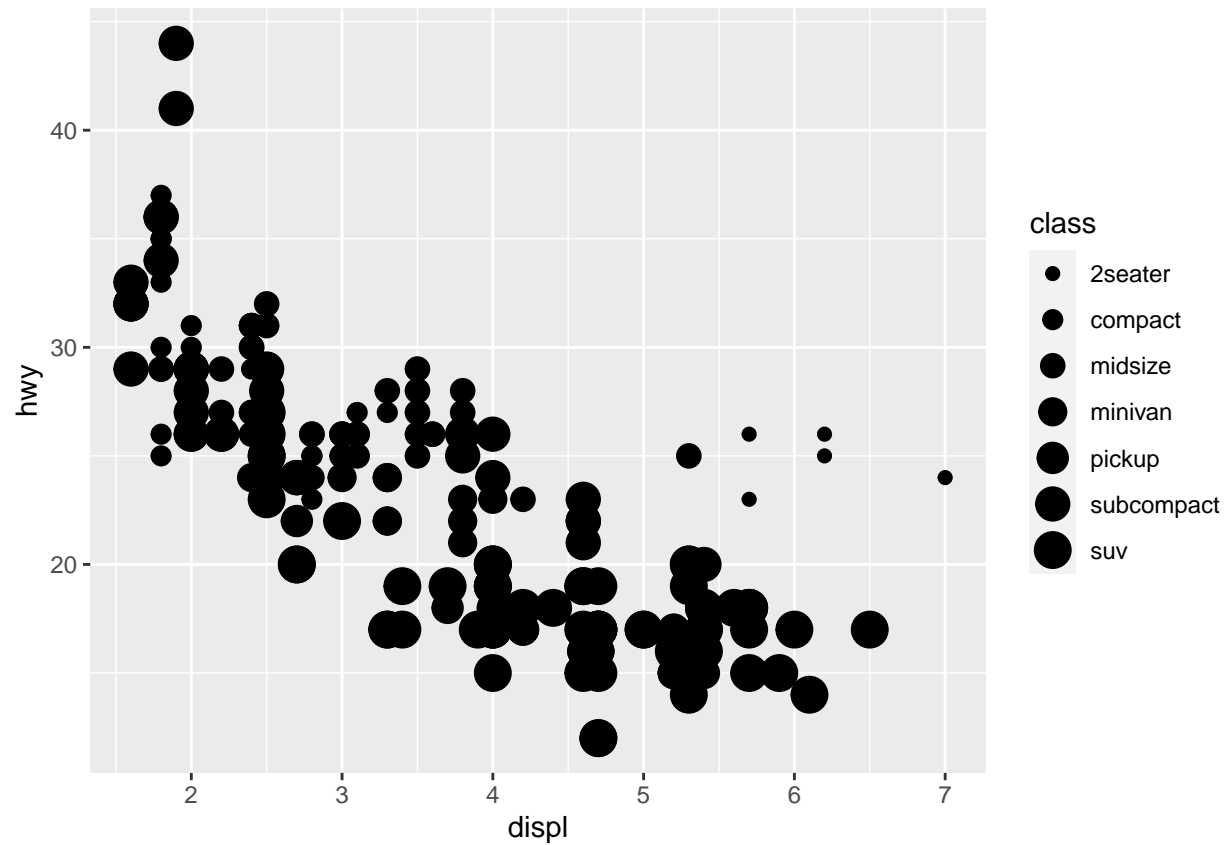
3.3

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy, colour = class))
```



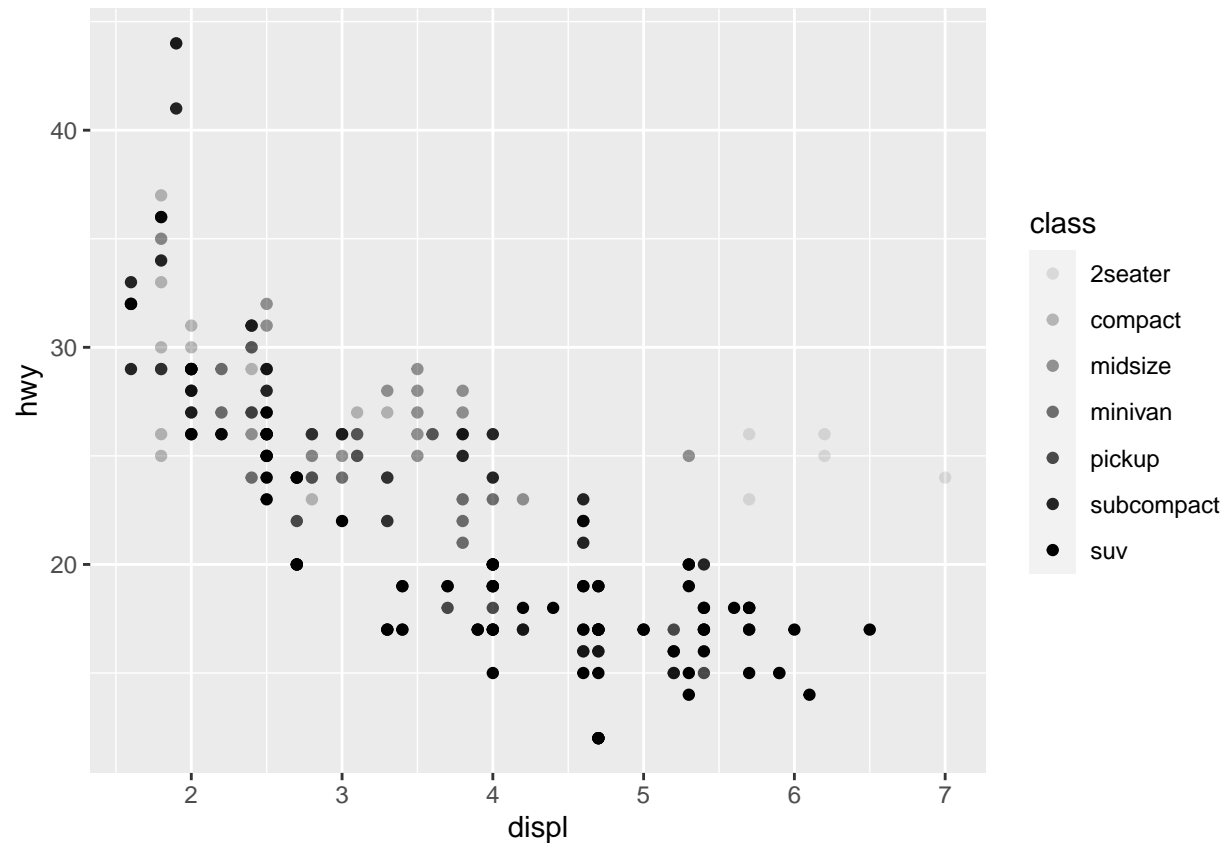
```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy, size = class))
```

```
## Warning: Using size for a discrete variable is not advised.
```



```
# Left  
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy, alpha = class))
```

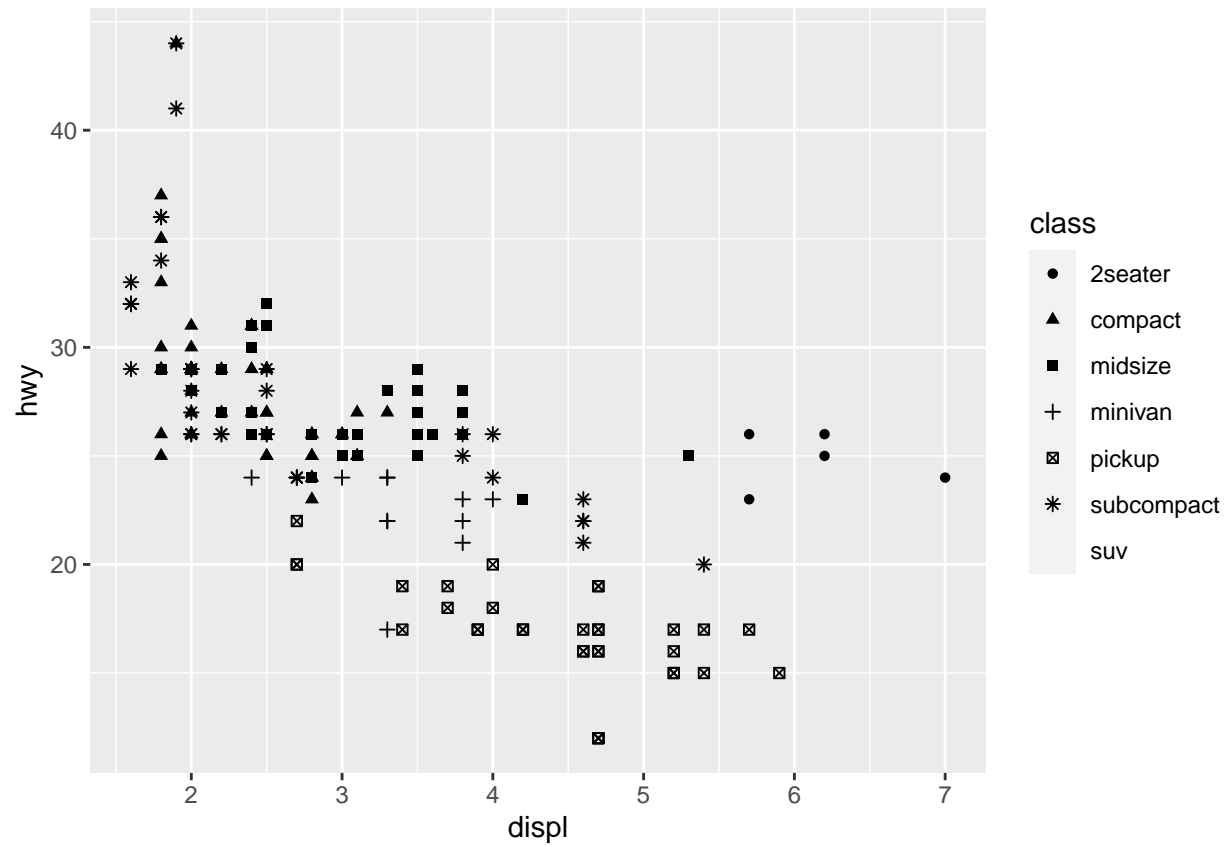
```
## Warning: Using alpha for a discrete variable is not advised.
```



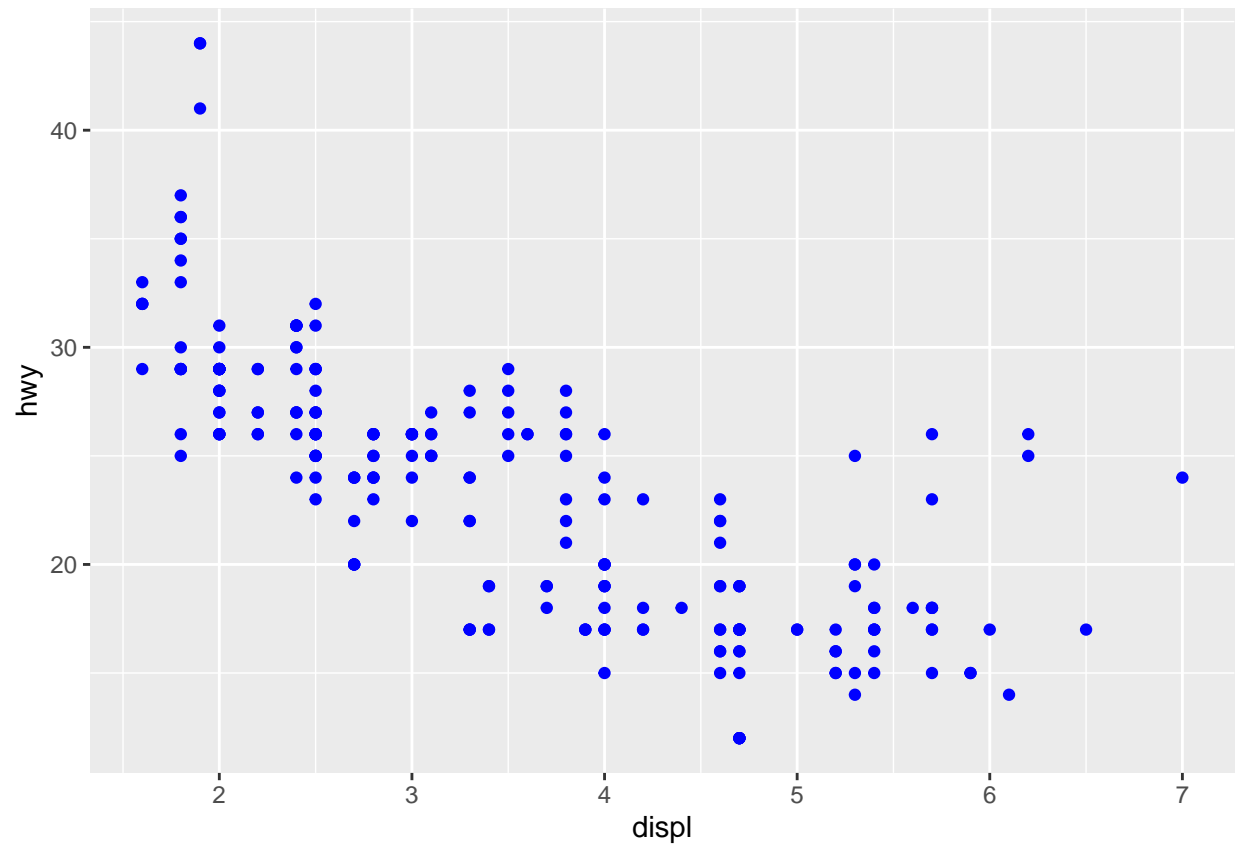
```
# Right
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, shape = class))
```

```
## Warning: The shape palette can deal with a maximum of 6 discrete values because
## more than 6 becomes difficult to discriminate; you have 7. Consider
## specifying shapes manually if you must have them.
```

```
## Warning: Removed 62 rows containing missing values (geom_point).
```

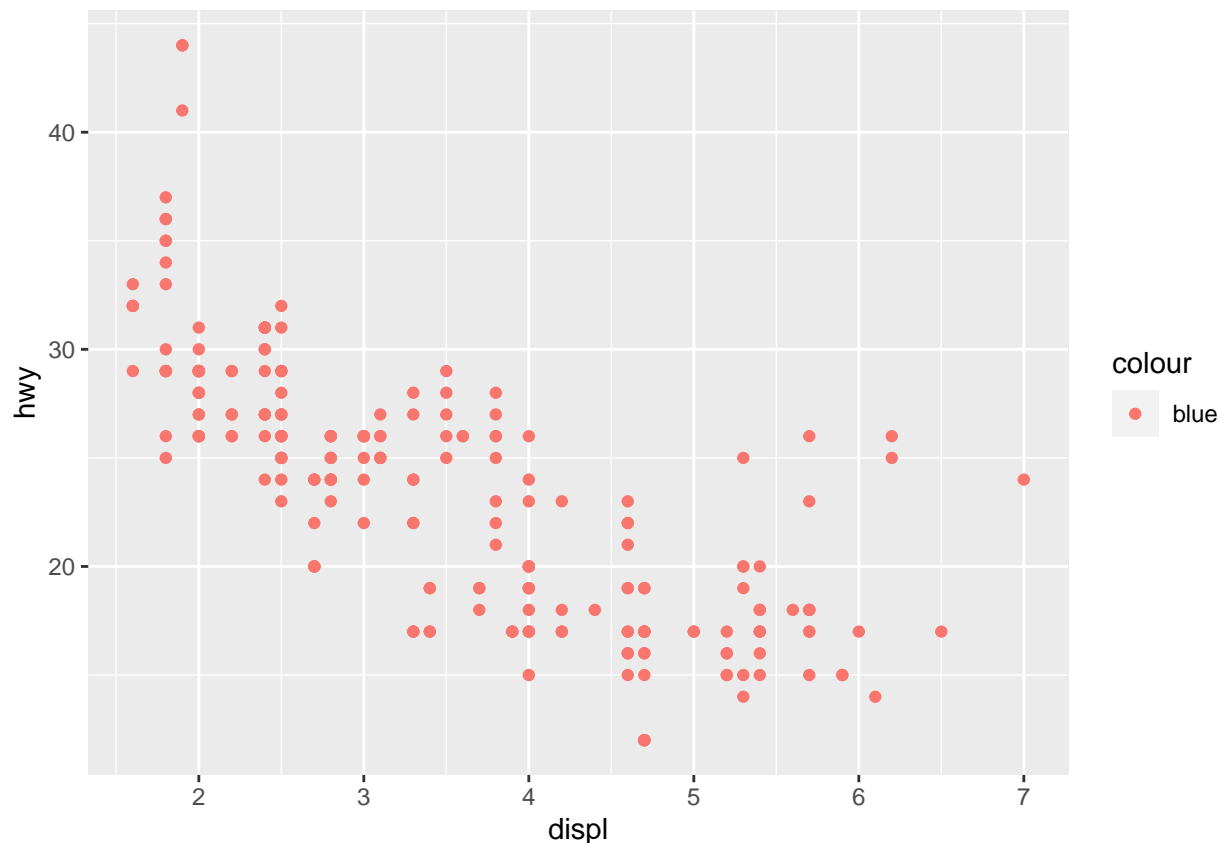
```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy), color = "blue")
```



3.3.1

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

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



The colour is not blue because “color = ‘blue’ ” is inside of `aes()` when it needs to be outside. Right now it’s being read as a variable.

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`?

```
?mpg
```

Categorical: manufacturer, displ, model, cyl, trans, drv, fl, class

Continious: year, cty, hwy

```
mpg
```

```
## # A tibble: 234 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~ f      18    29 p    comp~
## 2 audi         a4        1.8  1999    4 manu~ f      21    29 p    comp~
## 3 audi         a4        2    2008    4 manu~ f      20    31 p    comp~
## 4 audi         a4        2    2008    4 auto~ f      21    30 p    comp~
## 5 audi         a4        2.8  1999    6 auto~ f      16    26 p    comp~
## 6 audi         a4        2.8  1999    6 manu~ f      18    26 p    comp~
## 7 audi         a4        3.1  2008    6 auto~ f      18    27 p    comp~
```

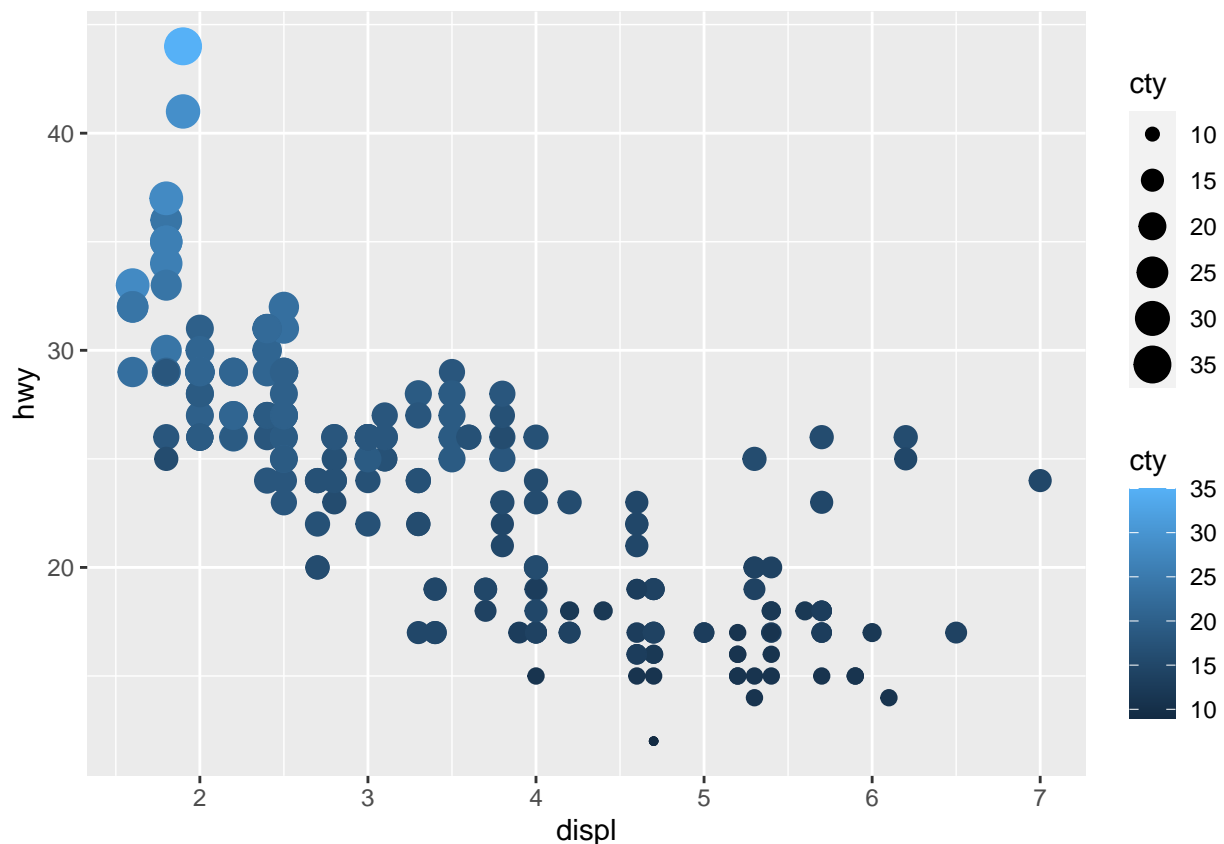
```
## 8 audi      a4 quattro  1.8 1999    4 manu~ 4      18   26 p   comp~
## 9 audi      a4 quattro  1.8 1999    4 auto~ 4      16   25 p   comp~
## 10 audi     a4 quattro  2    2008    4 manu~ 4      20   28 p   comp~
## # ... with 224 more rows
```

We can see this with mpg by simply looking at all the categories with limited values for answers. Something like year is continuous because they can make a new car every year while drv is categorical because there are only three types of drive trains.

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

```
# ggplot(data = mpg) +
#   geom_point(mapping = aes(x = displ, y = hwy, #colour = cty, size = cty, shape = cty))

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



A continuous variable cannot be mapped to shape where a categorical variable can.

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

As we can see above, when a variable is mapped to more than one aesthetic the legend will show both and the graph will take a combination of the two aesthetics.

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

```
?geom_point
```

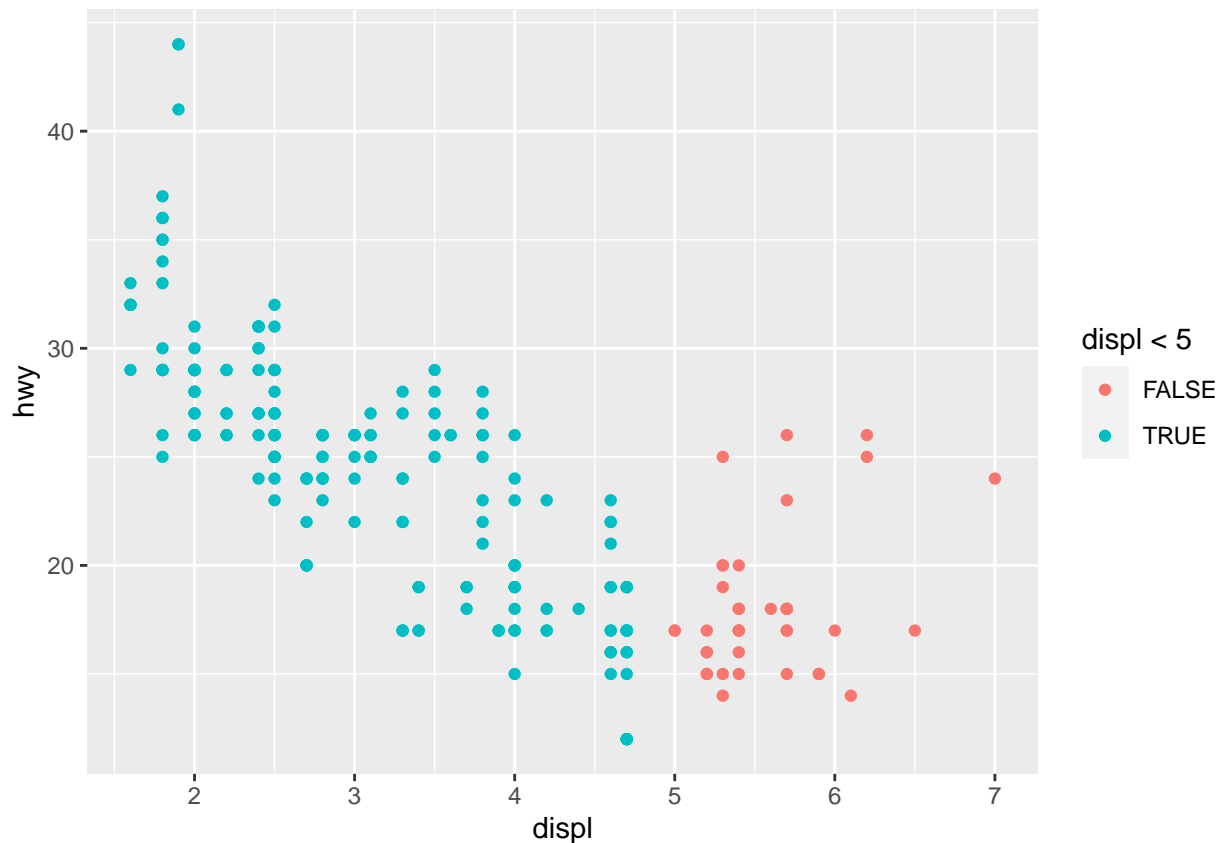
```
vignette("ggplot2-specs")
```

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

Stroke changes the thickness of a border. The size and color of it can be changed.

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.

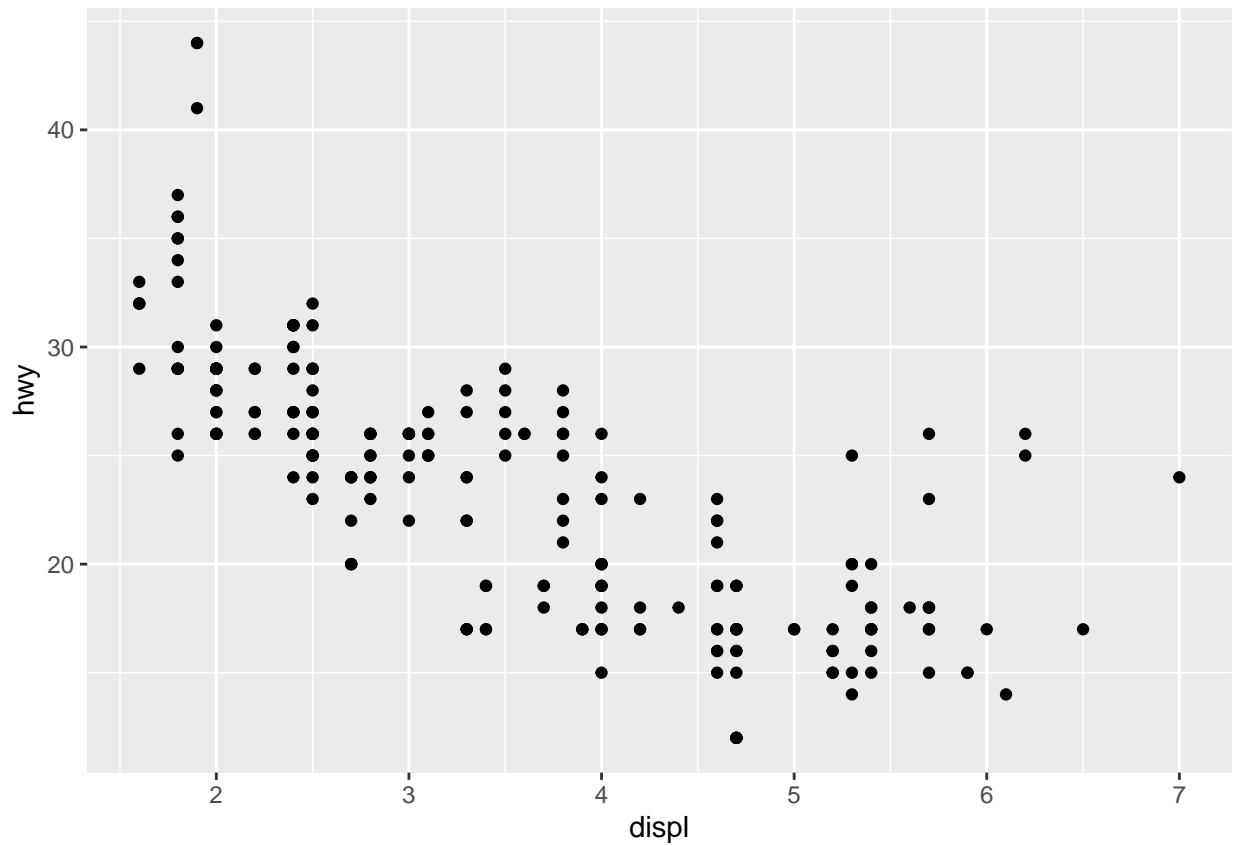
```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy, colour = displ < 5))
```



In this instance doing `displ < 5` makes a true/false colour scheme based on if the value of displ is less than 5 or greater than or equal to 5.

3.4

```
#ggplot(data = mpg)  
#+ geom_point(mapping = aes(x = displ, y = hwy))  
  
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy))
```



?function__name