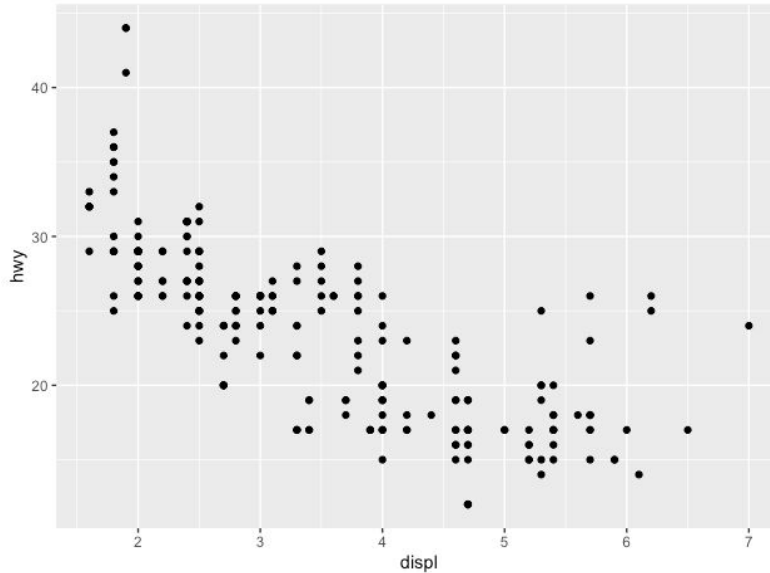


Exercise 1:

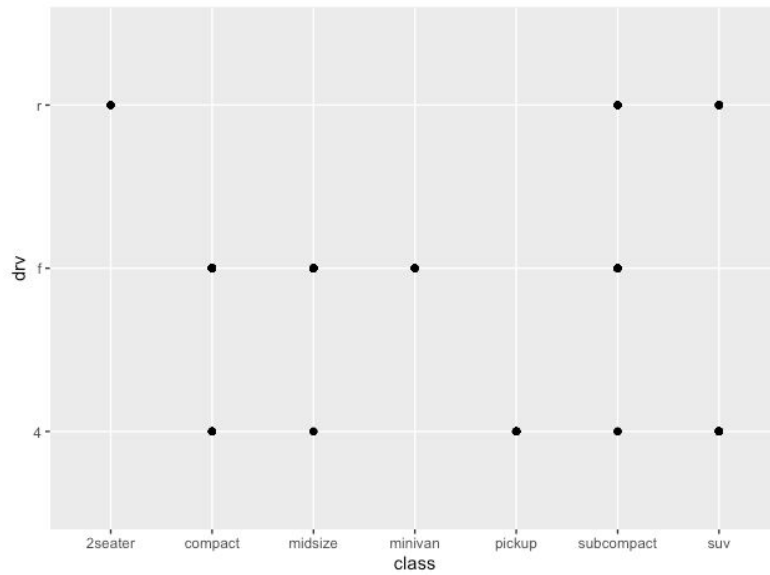
```
p <- ggplot(mpg, aes(displ, hwy))  
p + geom_point()
```



This plot does confirm the relationship we expected, as it shows a higher engine displacement usually pertains to lower highway mpg. The outliers in this data are possibly due to cylinder shut-off features in many modern V8s that turn off half of the cylinders in the engine on the highway. These outliers may also be due to hybrid electric vehicles or diesel engines.

If we run this same code on *class* and *drv*, however, the visualization is pretty meaningless:

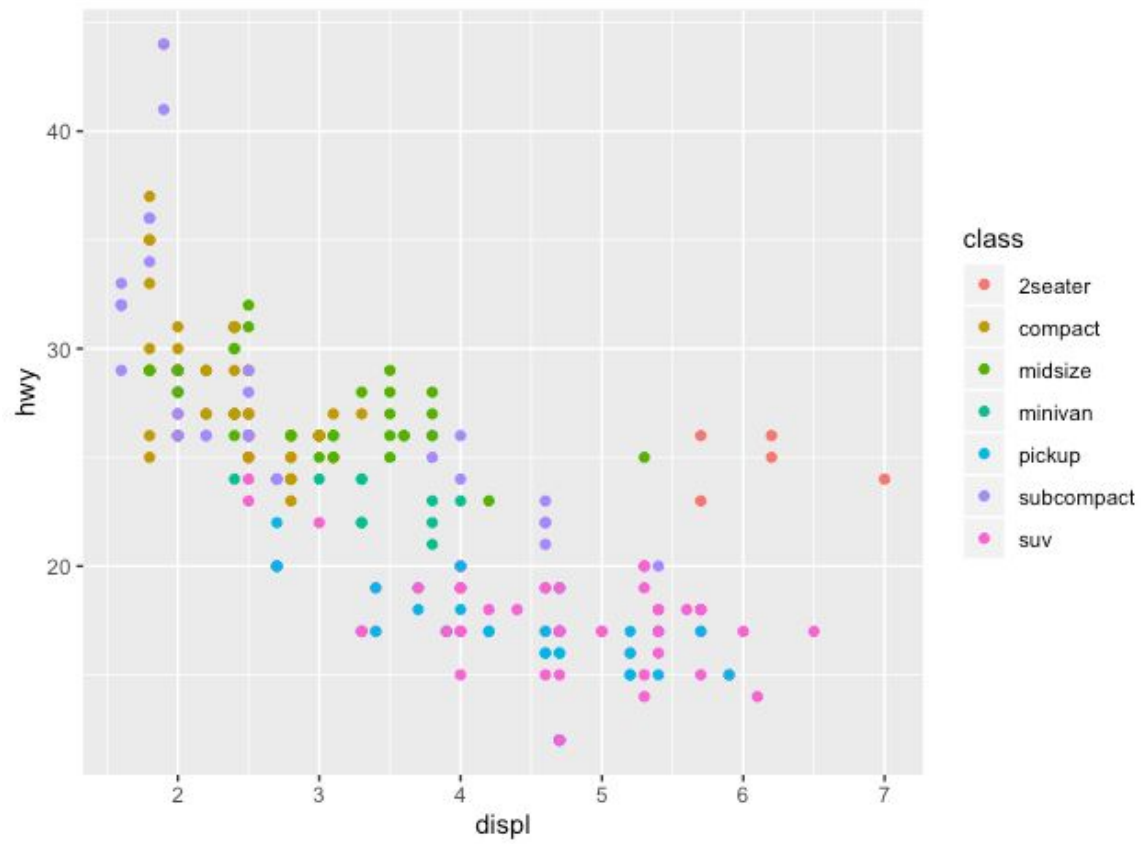
```
p <- ggplot(mpg, aes(class, drv))  
p + geom_point()
```



It's just not possible to draw any conclusions from this chart.

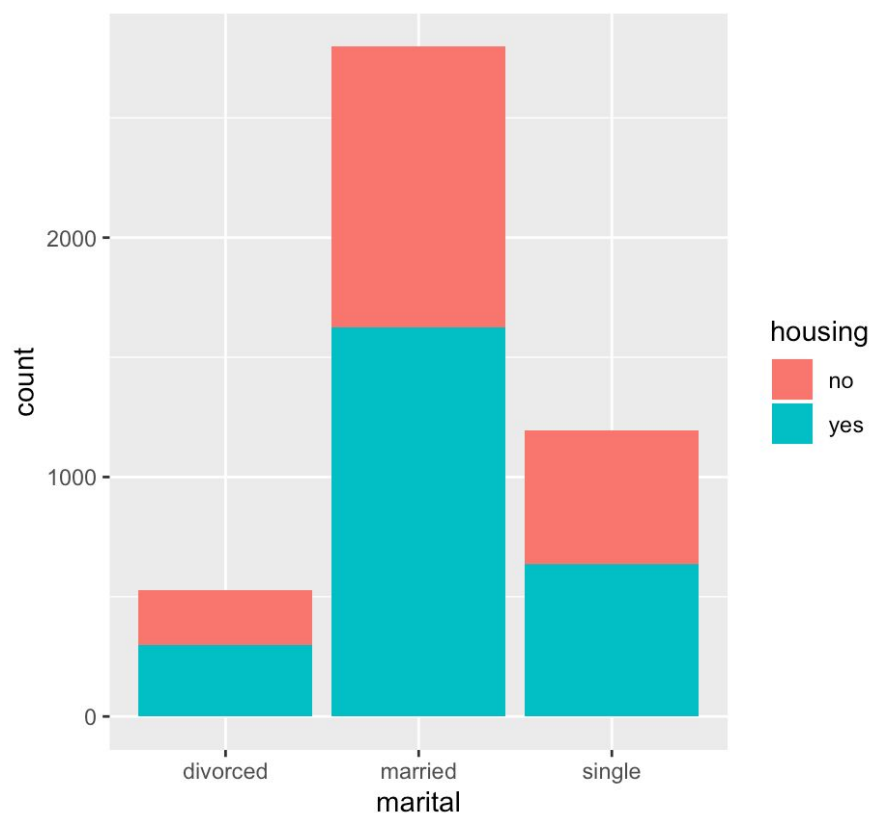
Exercise 1b:

```
p <- ggplot(mpg, aes(displ, hwy, color = class))
p + geom_point()
```



From this data it's easy to see that the same types of cars usually cluster in an area of similar engine displacements and highway fuel efficiency. For example, most of the compact cars have an engine displacement of under 3 liters and an average highway mpg of above 25, whereas most of the SUVs come in at over 4 liters and an average highway mpg of under 20.

Exercise 2:



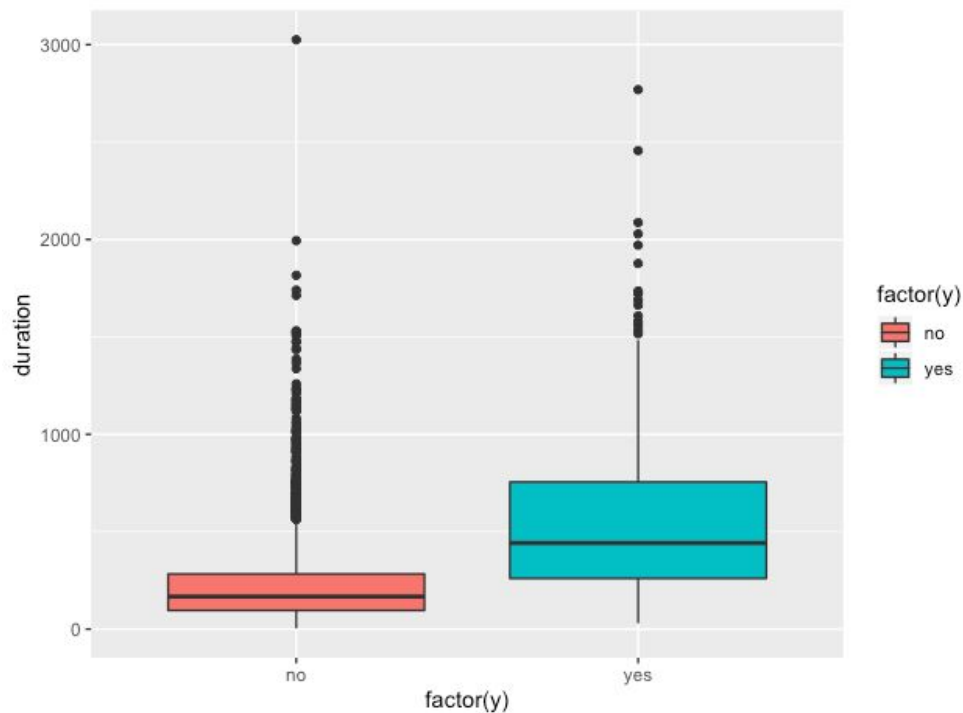
```
bankdata <- read.csv("bank.csv")
```

```
ggplot(data = bankdata, aes(x = marital, fill = housing)) +  
  geom_bar()
```

One of the visualizations that we created was the relationship between marital status and housing loans. We hypothesized that couples who were married may be more likely than single people to want a mortgage. We believed this because it could be possible

that single people are more willing to rent housing than to buy a house. When we created the visualizations, you can see that there are, in fact, more married people who have obtained these loans, but the percentage of married people getting mortgages is only a small amount larger than single people. For marketers, this could be useful because if they know that a lead is married or divorced, it may be easier to push for them to get a mortgage through this bank, while if they are single, it may be more useful to push for other types of loans.

```
bankData <- read.csv(file = 'bank.csv')  
p <- ggplot(bankData, aes(factor(y), duration))  
p + geom_boxplot(aes(fill = factor(y)))
```



The other relationship that we examined was the duration of the phone call and whether or not the person is subscribed to term deposits. We believed that if a phone call was longer, this may mean that the marketers took more time to establish a relationship with the customer, and in turn, were able to make a deal. With the visualizations, we were able to see that this was true. From a marketing perspective, if a marketer were to take some extra time, it may be easier for them to make a deal. On the other hand, this could also be true because if someone is signing up for the term deposit, their phone call would be longer since it would take longer to set everything up.