Assignment 3

Neev Shaw

September 17, 2025

3.2.4 Exercises

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

[1]:
```
library(ggplot2)
ggplot(data=mpg)
```

We see a blank gray screen that is the background for any ggplot. We haven't told it what to draw - just to initialize a graph - so it's just a gray box!

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

[2]:
```
print(dim(mpg))
```

```
[1] 234  11
```

There are 234 rows and 11 columns.

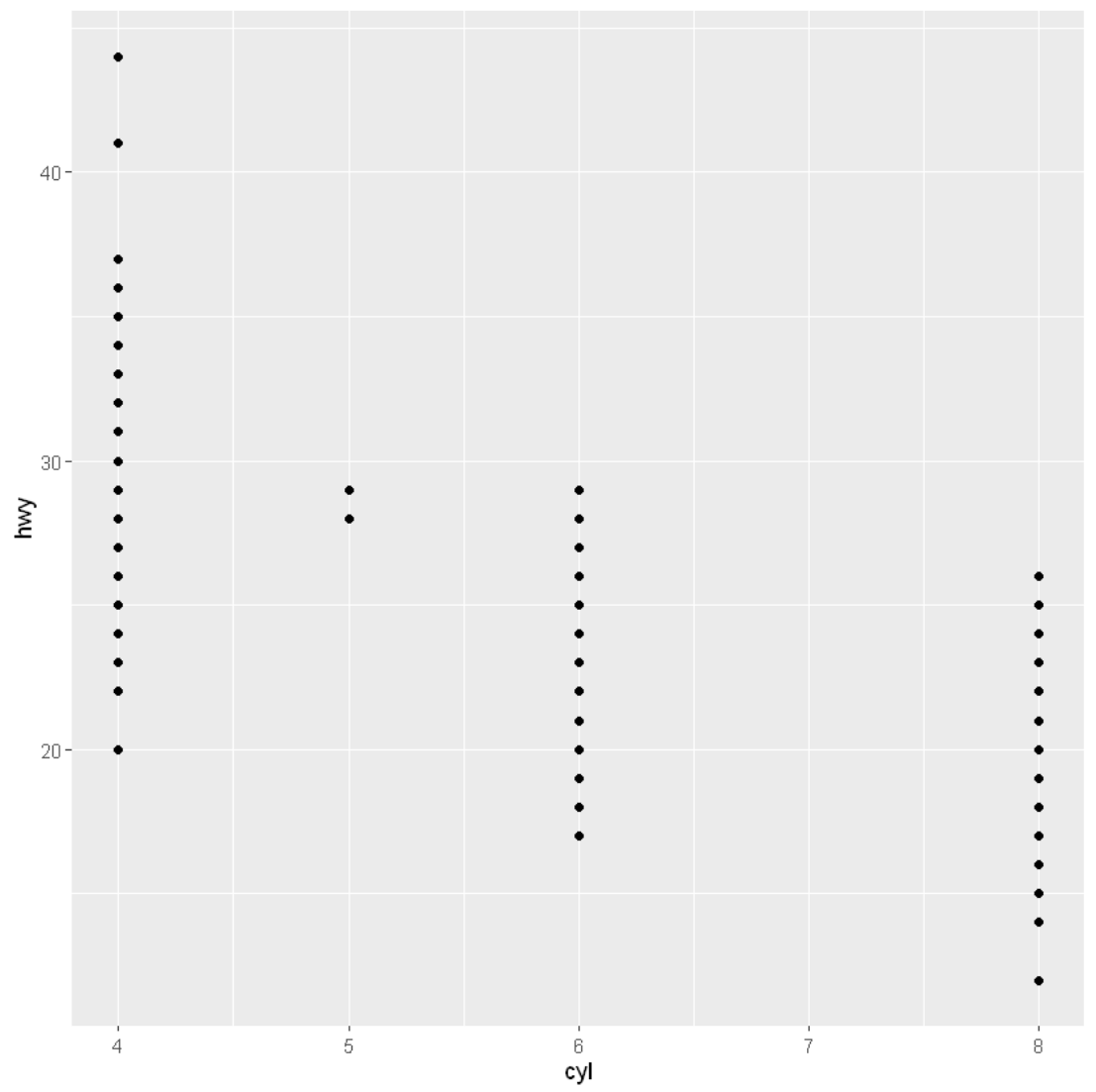
3. What does the `drv` variable describe? Read the help for `?mpg` to find out.

[ ]:
```
?mpg # Output hidden, because it is extremely big.
```

`drv` represents the type of drive train (f = front-wheel drive, r = rear-wheel drive, 4 = 4-wheel drive)

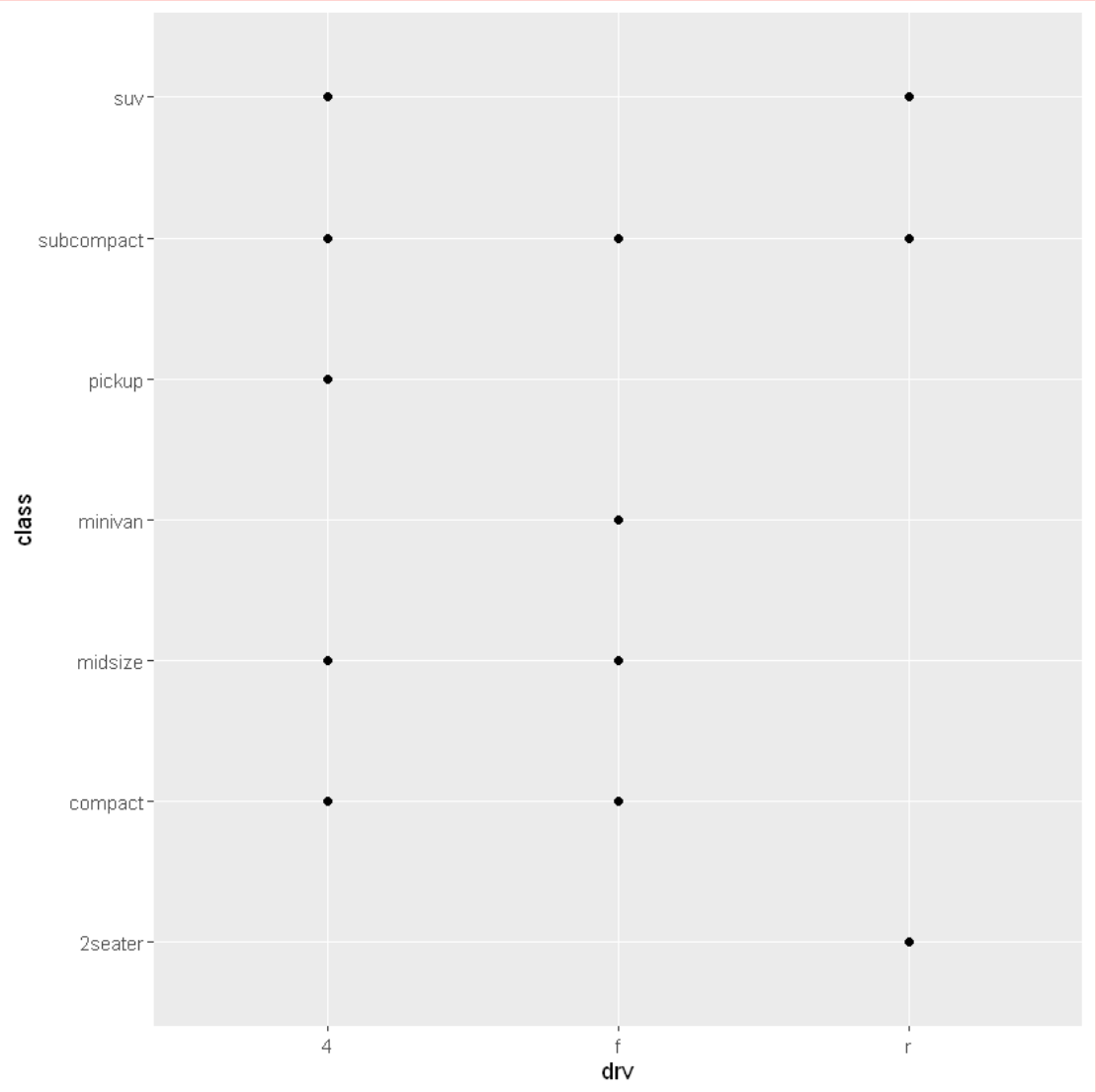
4. Make a scatterplot of `hwy` vs `cyl`.

[4]:
```
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?

[5]:
```
ggplot(data=mpg) + geom_point(mapping=aes(x=drv, y=class))
```



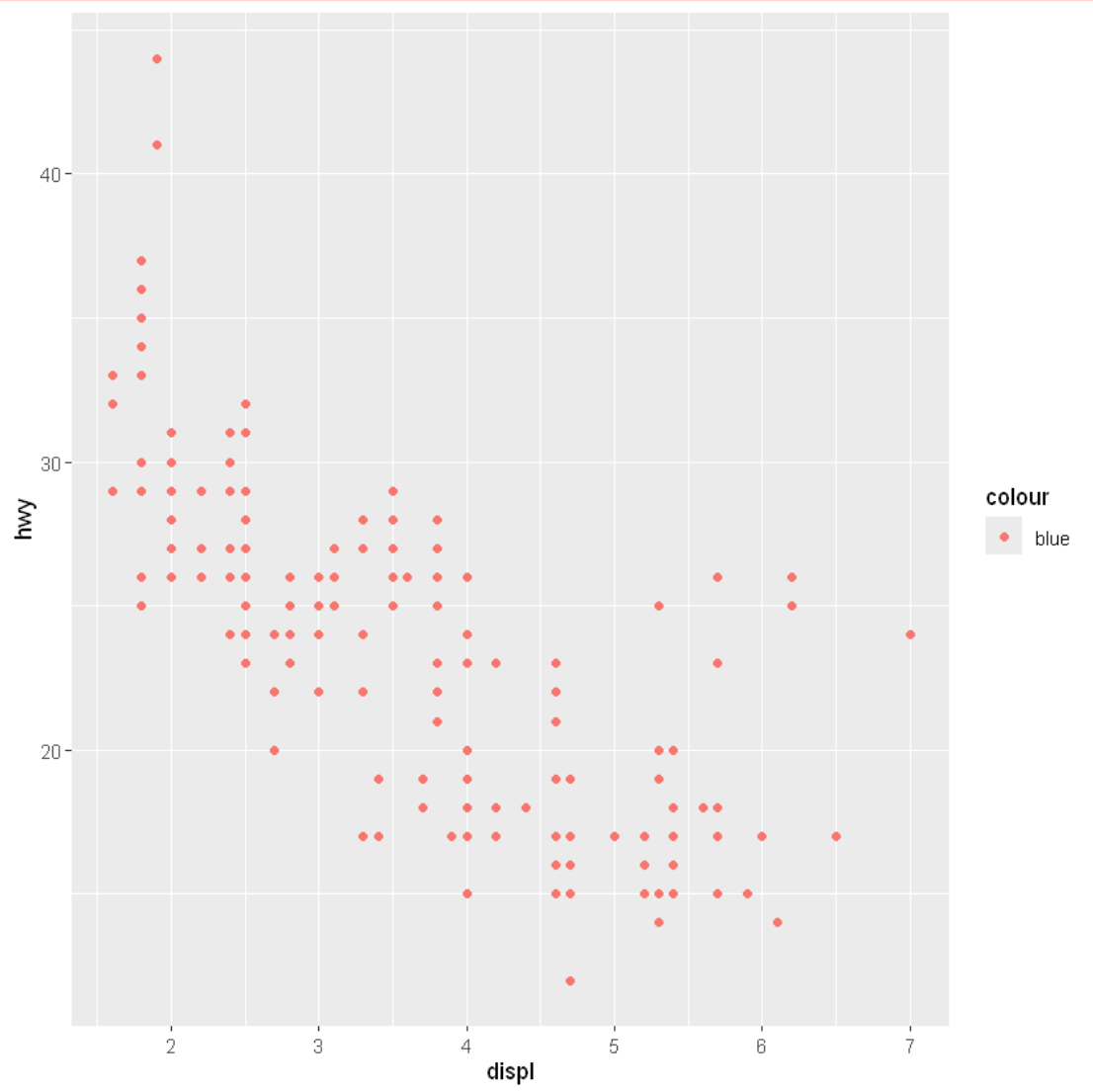
This plot is not useful because both of the variables are categorical - which means if a car has a particular `drv` and `class`, it will just appear as a point corresponding to that pair. If there are multiple cars at a point, we would never know because the points overlap. This means we can't even analyze the distriution at a particular value of `class` or `drv`, since we have no idea how many points are at each coordinate pair. This is a pretty useless graph!

3.3.1 Exercises

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

Original code

[6]:
```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = "blue"))
```



Fixed code:

[7]:
```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy), color = "blue")
```

Color is not supposed to change based on the variables, so it doesn't go inside the `aes` function. Instead it is set as a global property outside the `aes` (but still within the `geom_point`) that changes the color of *all* the points.