01-006

Alex Cookson

2020-07-10

```
library(tidyverse)
```

Question 1

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

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


A blank gray plot. We don't see anything because we haven't set the `aes()` or `geom()` layers, which tell us *what* to graph and *how* to graph the data we've specified.

Question 2

How many rows are in `mtcars`? How many columns?

The `glimpse()` function gives an overview of the data, including the number of rows and columns:

```
glimpse(mtcars)
```

```
## Rows: 32
## Columns: 11
## $ mpg  <dbl> 21.0, 21.0, 22.8, 21.4, 18.7, 18.1, 14.3, 24.4, 22.8, 19.2, 17...
## $ cyl  <dbl> 6, 6, 4, 6, 8, 6, 8, 4, 4, 6, 6, 8, 8, 8, 8, 8, 8, 4, 4, 4, 4,...
## $ disp <dbl> 160.0, 160.0, 108.0, 258.0, 360.0, 225.0, 360.0, 146.7, 140.8,...
## $ hp   <dbl> 110, 110, 93, 110, 175, 105, 245, 62, 95, 123, 123, 180, 180, ...
## $ drat <dbl> 3.90, 3.90, 3.85, 3.08, 3.15, 2.76, 3.21, 3.69, 3.92, 3.92, 3....
## $ wt   <dbl> 2.620, 2.875, 2.320, 3.215, 3.440, 3.460, 3.570, 3.190, 3.150,...
## $ qsec <dbl> 16.46, 17.02, 18.61, 19.44, 17.02, 20.22, 15.84, 20.00, 22.90,...
## $ vs   <dbl> 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1,...
## $ am   <dbl> 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0,...
## $ gear <dbl> 4, 4, 4, 3, 3, 3, 3, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 4, 4, 4, 3,...
## $ carb <dbl> 4, 4, 1, 1, 2, 1, 4, 2, 2, 4, 4, 3, 3, 3, 4, 4, 4, 1, 2, 1, 1,...
```

`mtcars` has 32 rows and 11 columns.

If we were interested *only* in the number of rows and columns, we can use `dim()` (dimensions), which gives an vector with of number of rows and columns (rows is the first number).

```
dim(mtcars)
```

```
## [1] 32 11
```

Question 3

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

```
?mpg
```

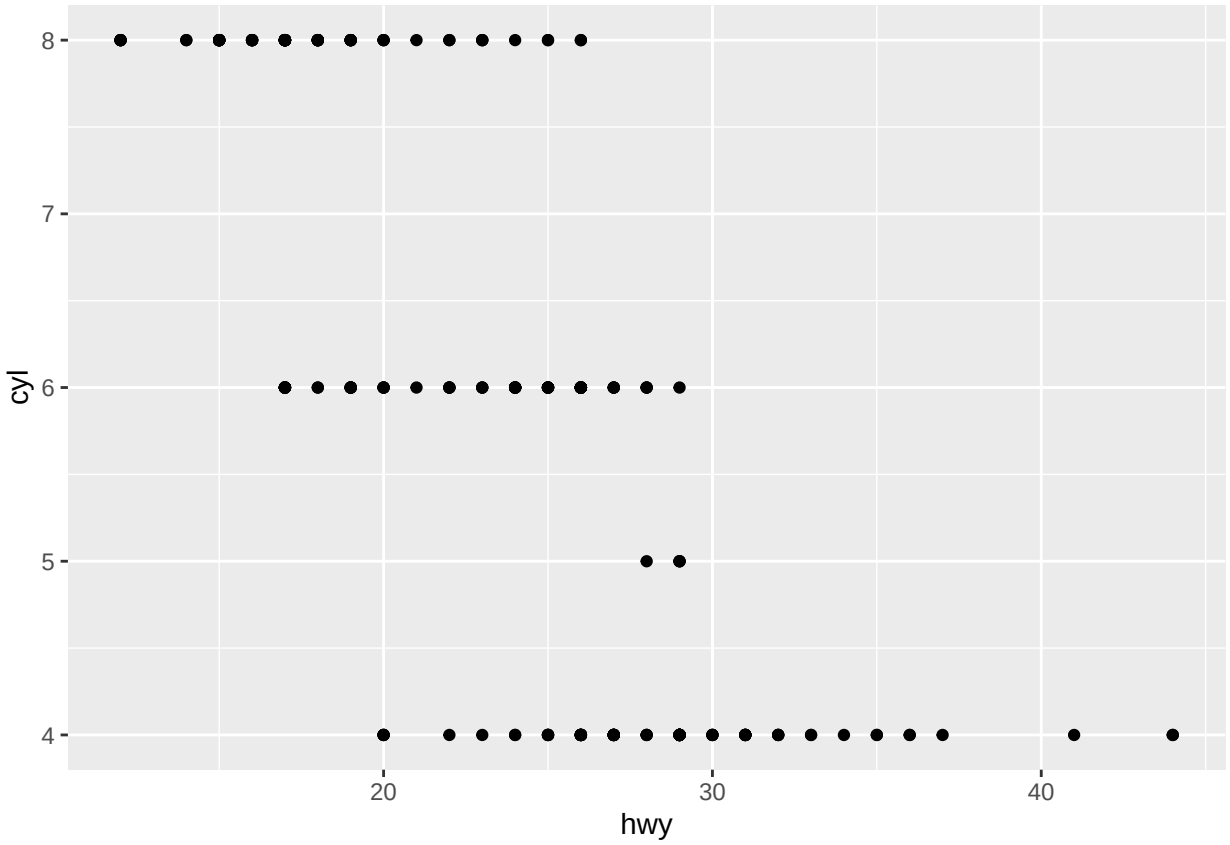
```
## starting httpd help server ... done
```

According the the documentation, `drv` is "the type of drive train, where f = front-wheel drive, r = rear wheel drive, 4 = 4wd"

Question 4

Make a scatterplot of `hwy` versus `cyl`.

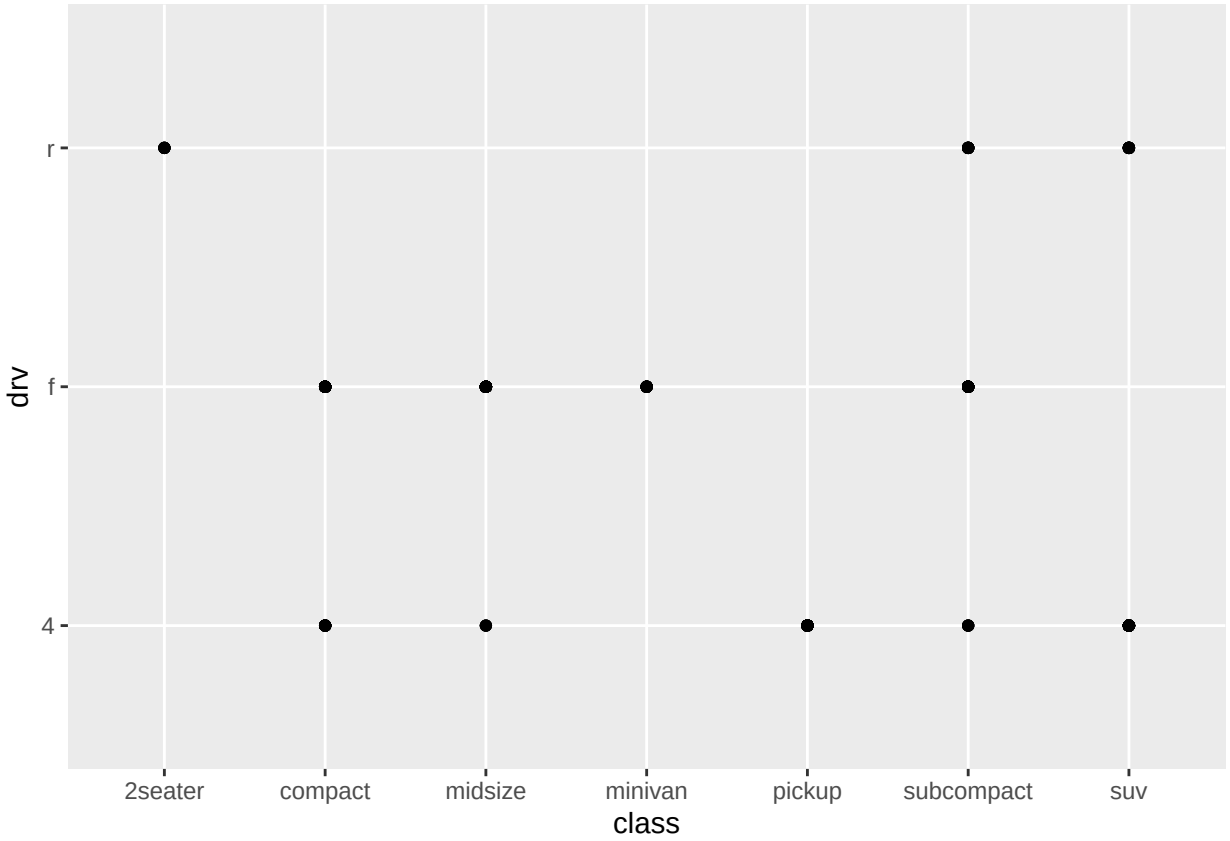
```r
ggplot(
  # Specify the data
  data = mpg,
  # Specify what we are graphing (hwy on x-axis, cyl on y-axis)
  mapping = aes(x = hwy, y = cyl)
) +
  # Specify how we want to graph it, a.k.a. the geometry (geom_point = scatterplot)
  geom_point()
```



Question 5

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

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



This isn't useful because there is no natural order to `class` or `drv`. Normally in a scatterplot, as we go farther right on the x-axis or up on the y-axis, a the *value* of what we're graphing increases. For example, in Question 4, `hwy` mileage increased as we went farther to the right.

With `class` and `drv`, we are dealing with **categorical** variables. `class` doesn't increase, it just changes. The same is true of `drv`. Rear-wheel drive isn't "more" or "less" than front-wheel drive; it's just different.

Scatterplots are suited to showuing the relationship between two variables as their values increase or decrease, but if you're dealing with variables that have no meaningful way of increasing or decreasing, a scatterplot becomes not useful.