01-015

Alex Cookson

2020-07-04

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

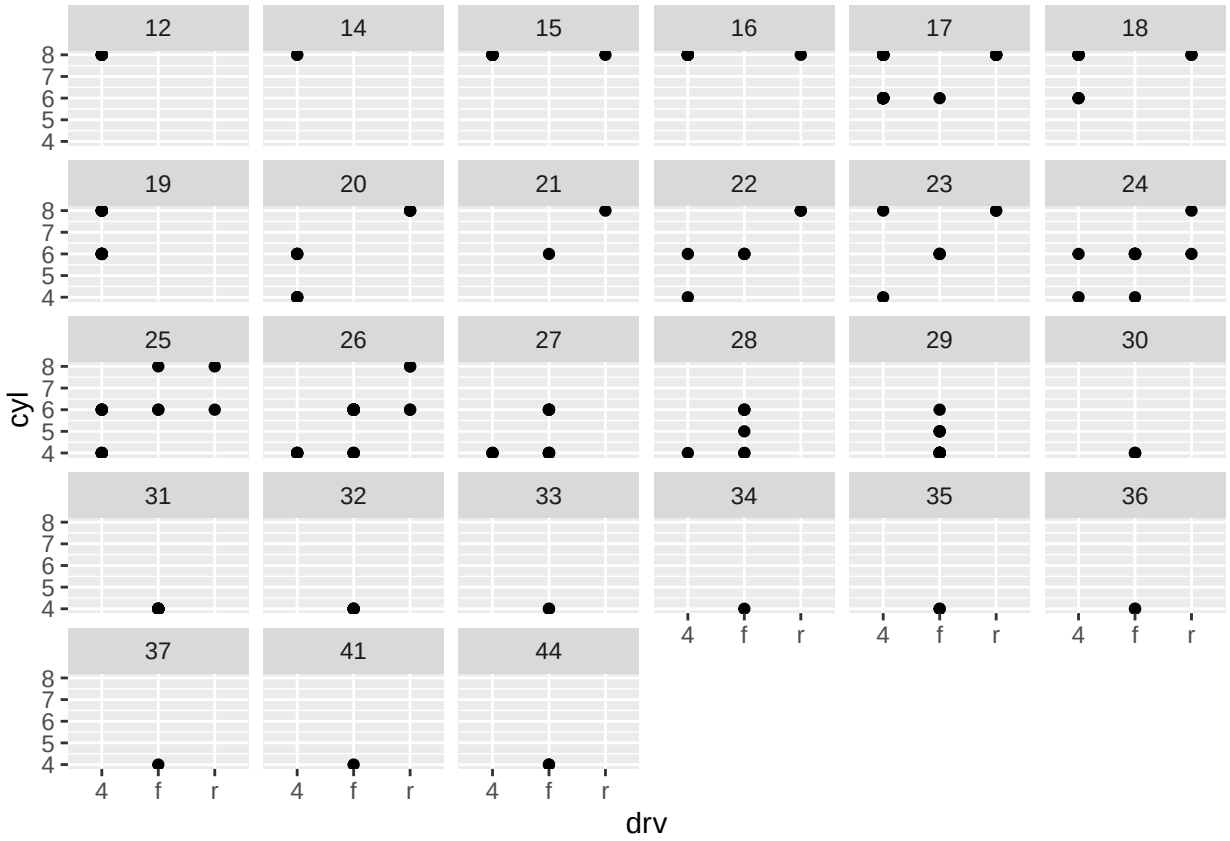
```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Question 1

What happens if you facet on a continuous variable?

Let's find out by graphing `drv` vs `cyl` and facetting by `hwy`:

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



ggplot converts the continuous variable into a categorical variable for the purposes of facetting, resulting in one facet per unique value in `hwy`.

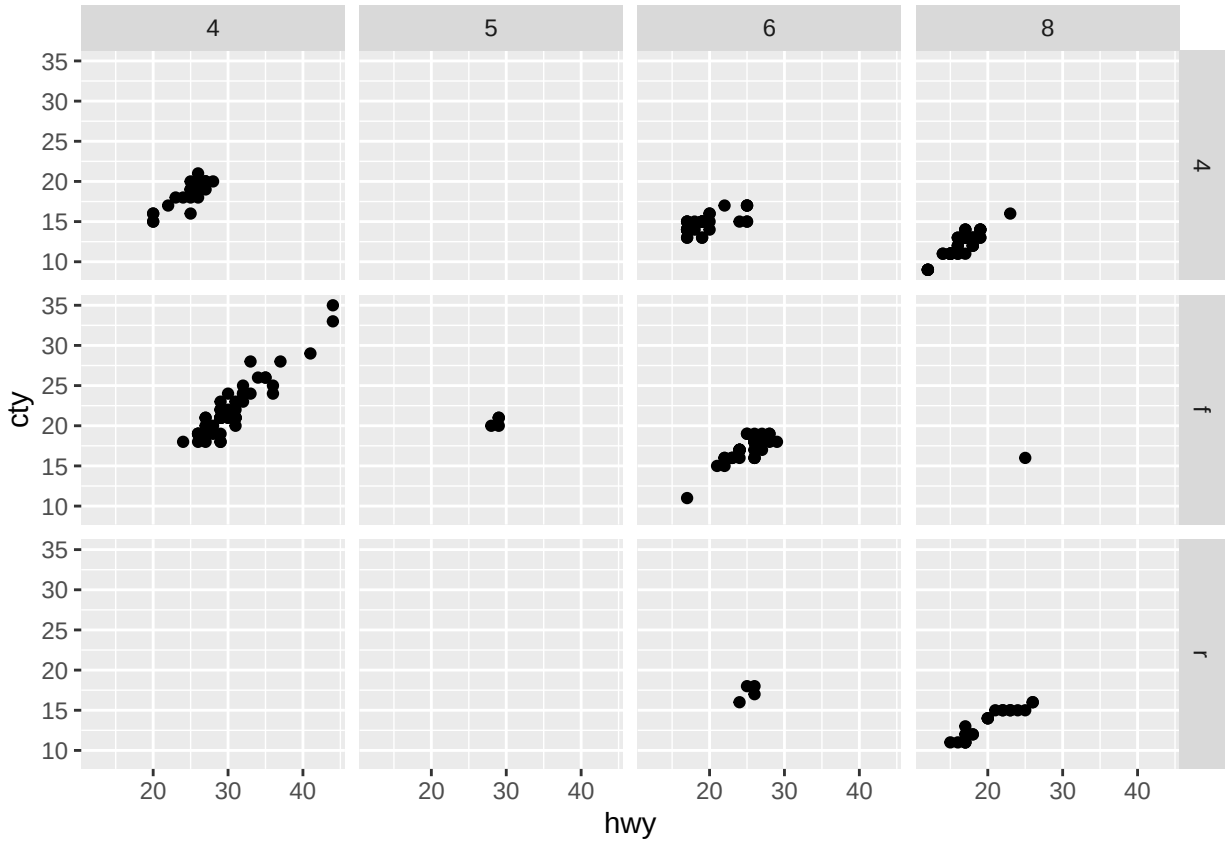
Question 2

What do the empty cells in a plot with `facet_grid(drv ~ cyl)` mean? How do they relate to this plot?

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

Let's start by actually graphing a plot with `facet_grid(drv ~ cyl)` to see what it looks like:

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

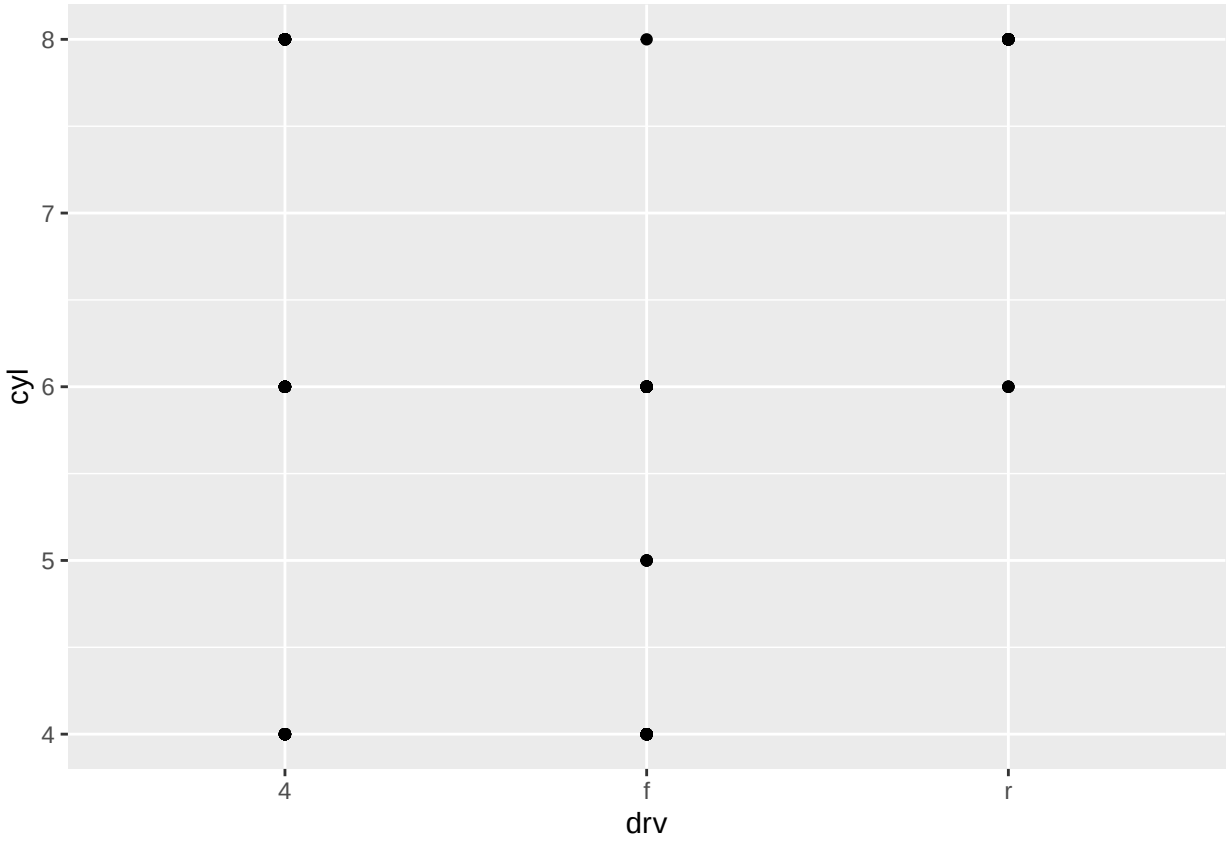


We have a facetted graph for each combination of `drv` (4, f, r) and `cyl` (4, 5, 6, 8) that has at least one observation. The points in the scatterplot represent the relationship between `hwy` and `cty` mileage for that combination of `drv` and `cyl` values.

If there are no points on the scatterplot, we can assume that there are no data points for that combination of `drv` and `cyl`. For example, there are no **5**-cylinder cars with **4**-wheel or **r**ear-wheel drive.

How does this relate to the plot given in the question?

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



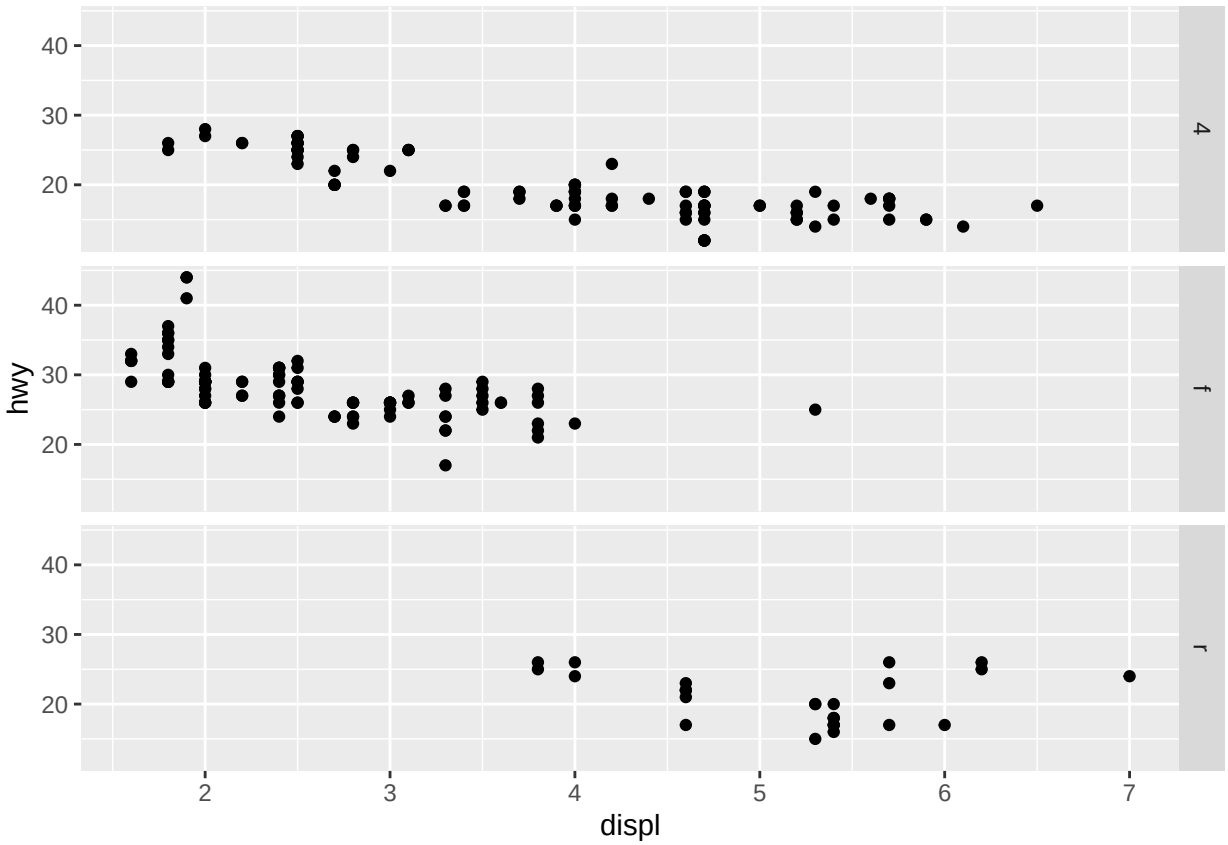
This plot shows the relationship between `drv` and `cyl`. Specifically, it shows which combinations of `drv` and `cyl` we have observations for. This is almost exactly what facetting does, except that instead of just showing a point (as in this plot), facetting shows us a whole scatterplot (as in the facetted plot above).

One difference is that this plot has not converted `cyl` to a categorical variable – it remains continuous. We can see that there is an axis tick for 7 cylinders, but no points associated with it, which means there are no 7-cylinder cars. In the facetted graph, because we had no 7-cylinder cars, **ggplot** didn't include 7 at all.

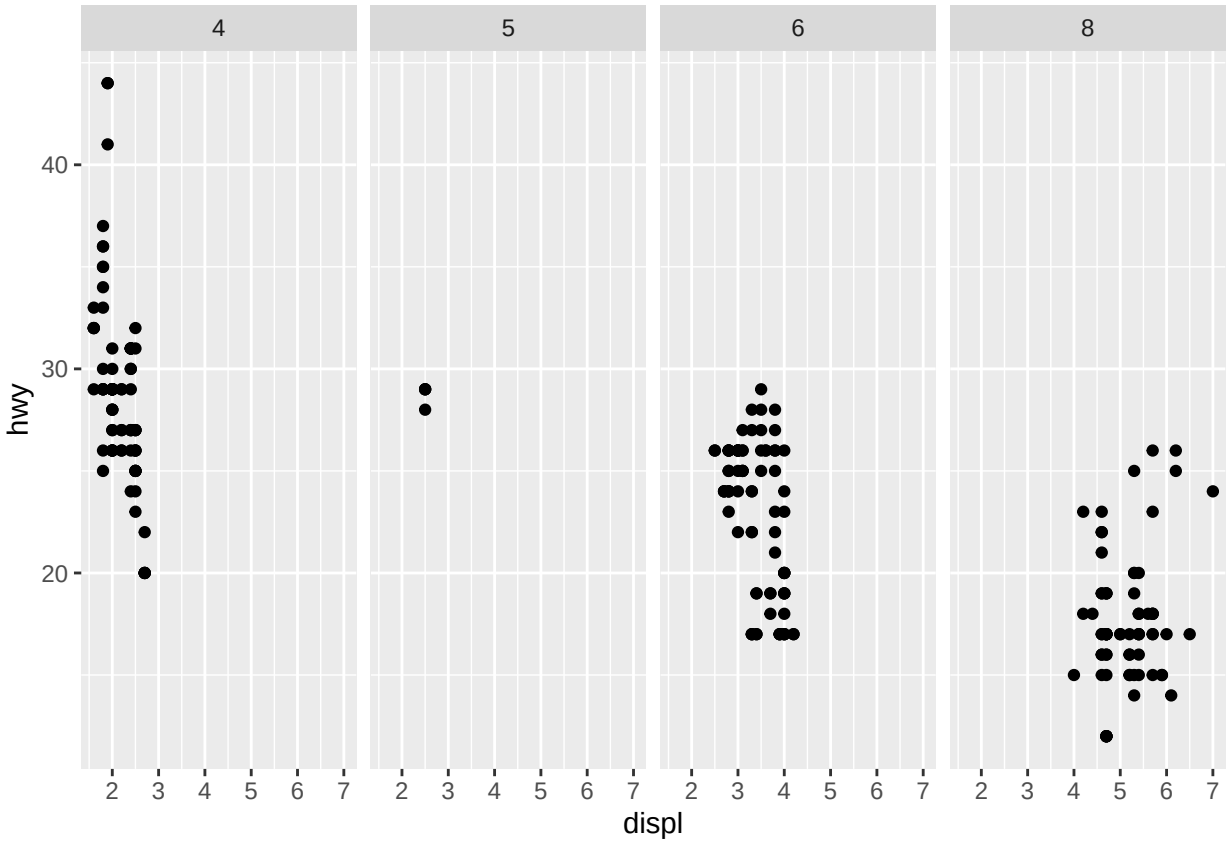
Question 3

What plots does the following code make? What does `.` do?

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




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



Question 4

Take the first faceted plot in this section:

```
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy)) +
  facet_wrap(~ class, nrow = 2)
```

What are the advantages to using faceting instead of the colour aesthetic? What are the disadvantages? How might the balance change if you had a larger dataset?

Question 5

Read `?facet_wrap`. What does `nrow` do? What does `ncol` do? What other options control the layout of the individual panels? Why doesn't `facet_grid()` have `nrow` and `ncol` variables?

Question 6

When using `facet_grid()` you should usually put the variable with more unique levels in the columns. Why?