

Lab 2

Joyce Gill

2026-01-22

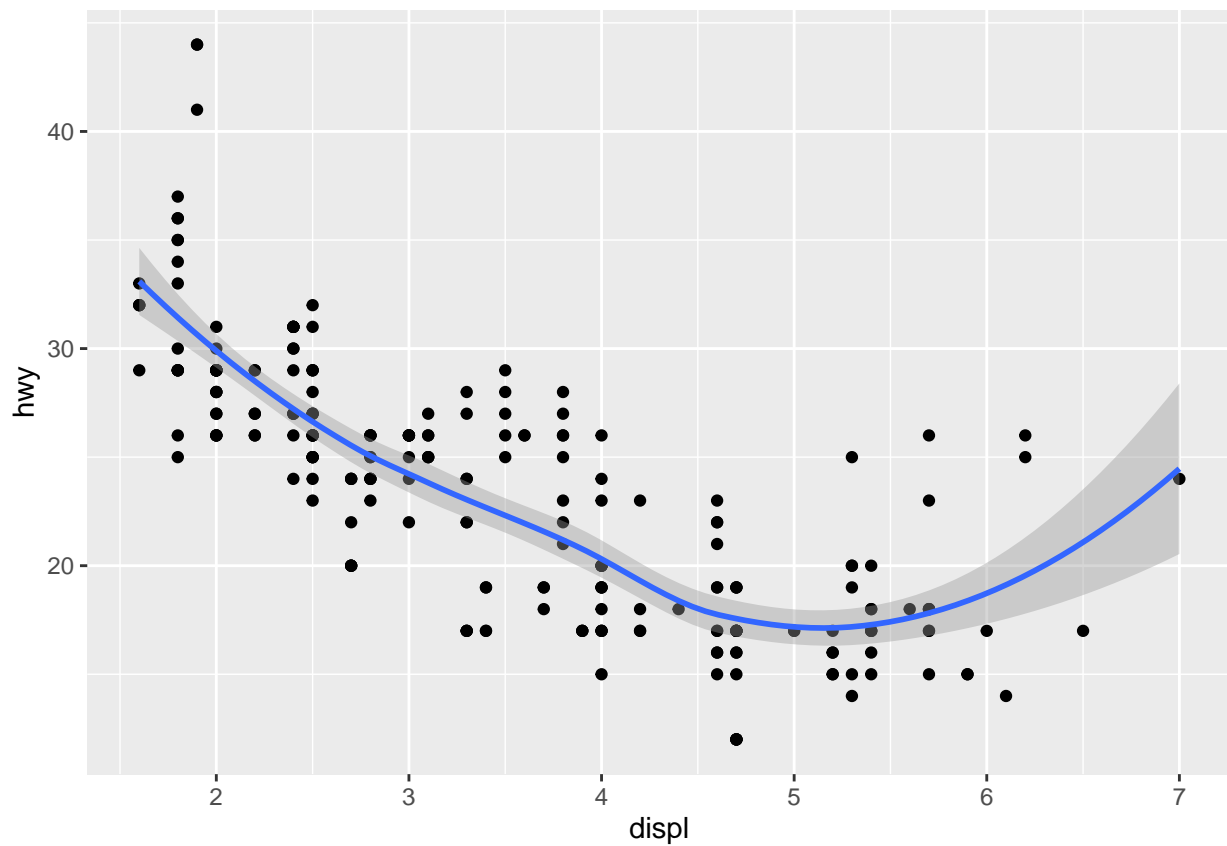
Question 1

```
library(ggplot2)
```

```
# Points then smooth (smooth drawn on top)
```

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

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

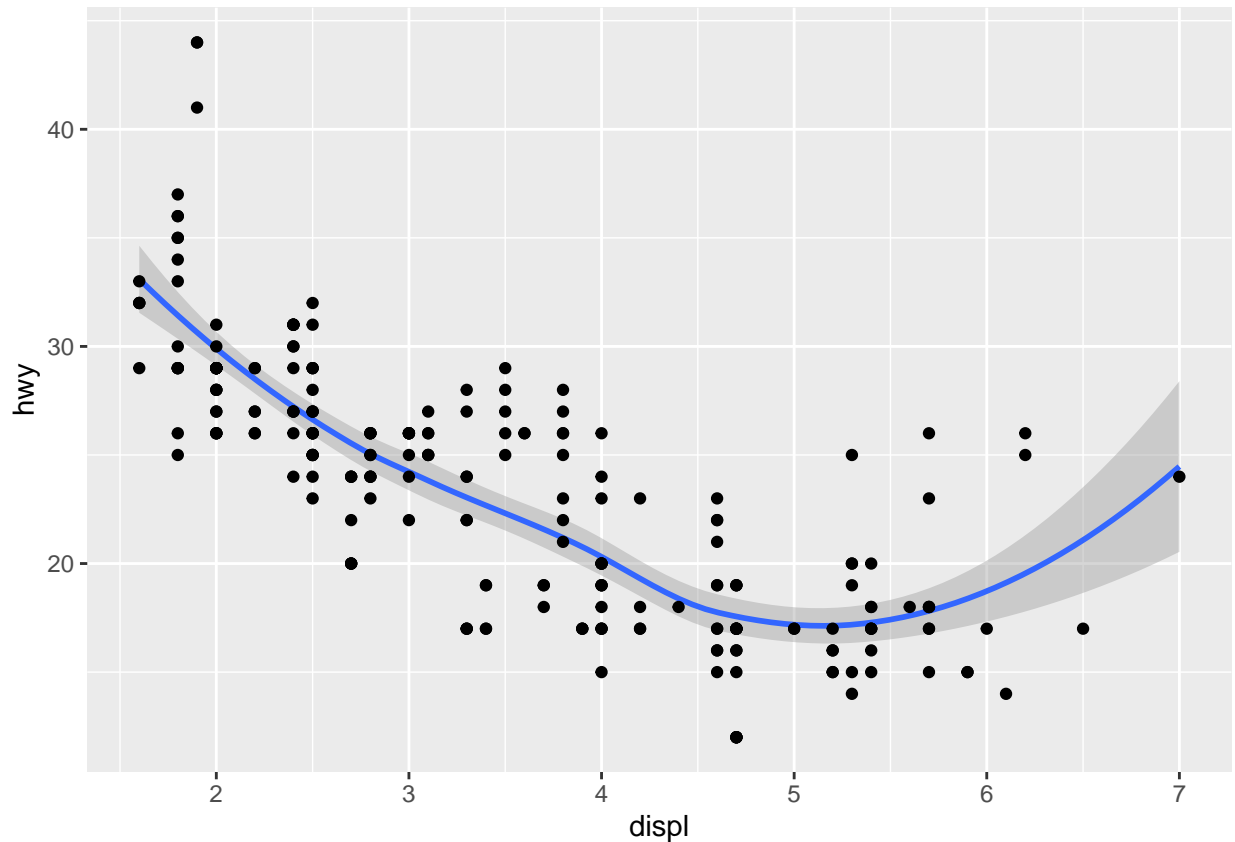


```
# Smooth then points (points drawn on top)
```

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

```
geom_point()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

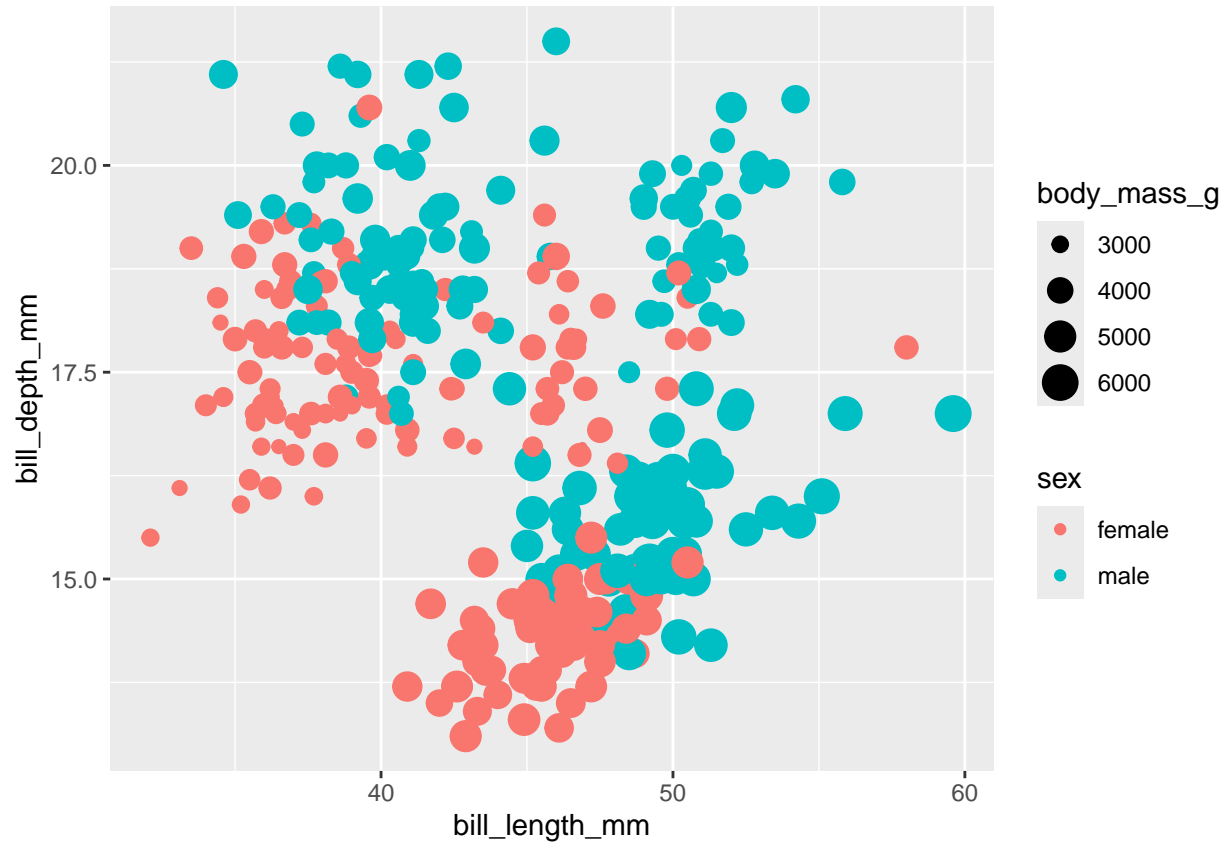


Layers are drawn in the order you add them. The later layer is plotted on top of earlier layers. So if you put `geom_smooth()` after `geom_point()`, the smooth line appears on top of the points; if you reverse them, many points will cover parts of the smooth.

Question 2

```
pengy <- read.csv("https://collinn.github.io/data/penguins.csv")

ggplot(pengy, aes(x = bill_length_mm, y = bill_depth_mm,
                  size = body_mass_g, color = sex)) +
  geom_point()
```



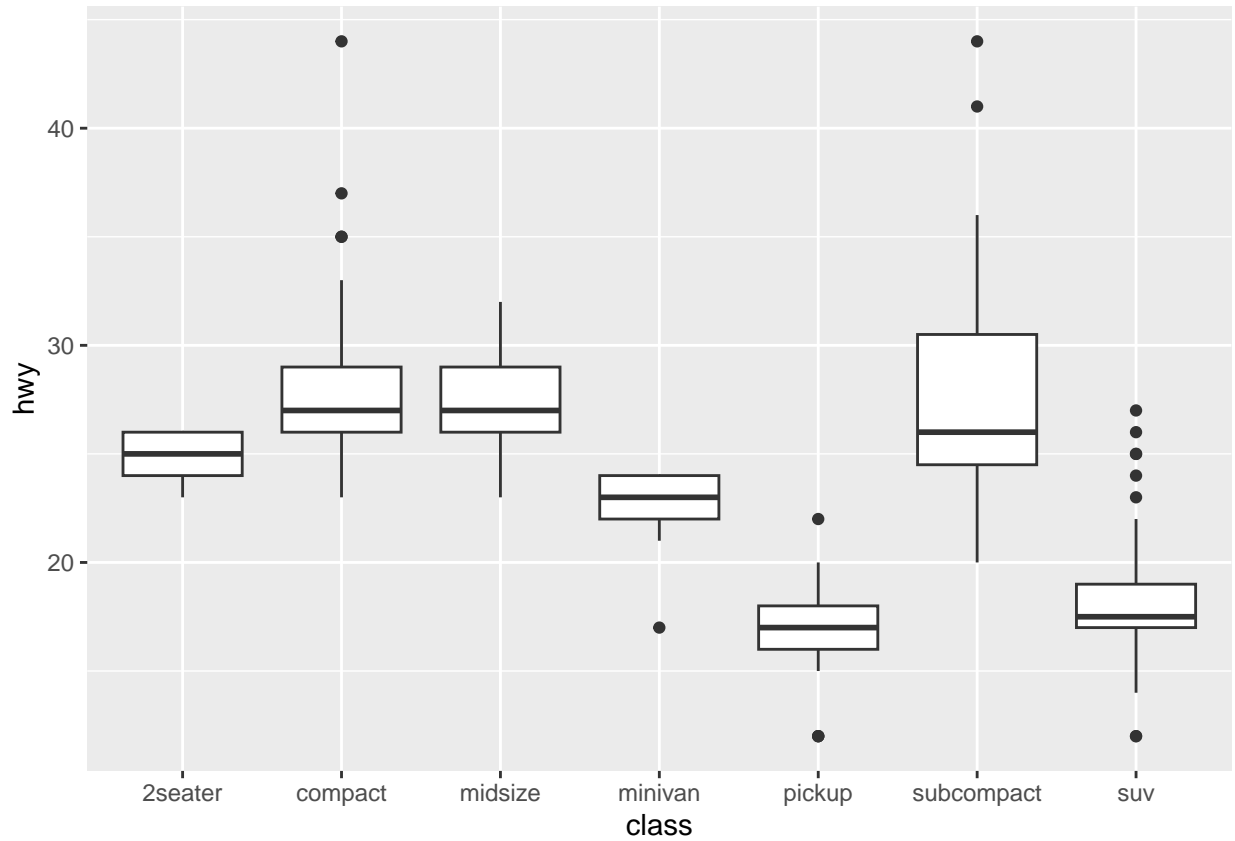
Data: pengy

Aesthetics: x = bill_length_mm, y = bill_depth_mm, size = body_mass_g, color = sex

Layers: geom_point()

Question 3

```
ggplot(mpg, aes(x = class, y = hwy)) +  
  geom_boxplot()
```

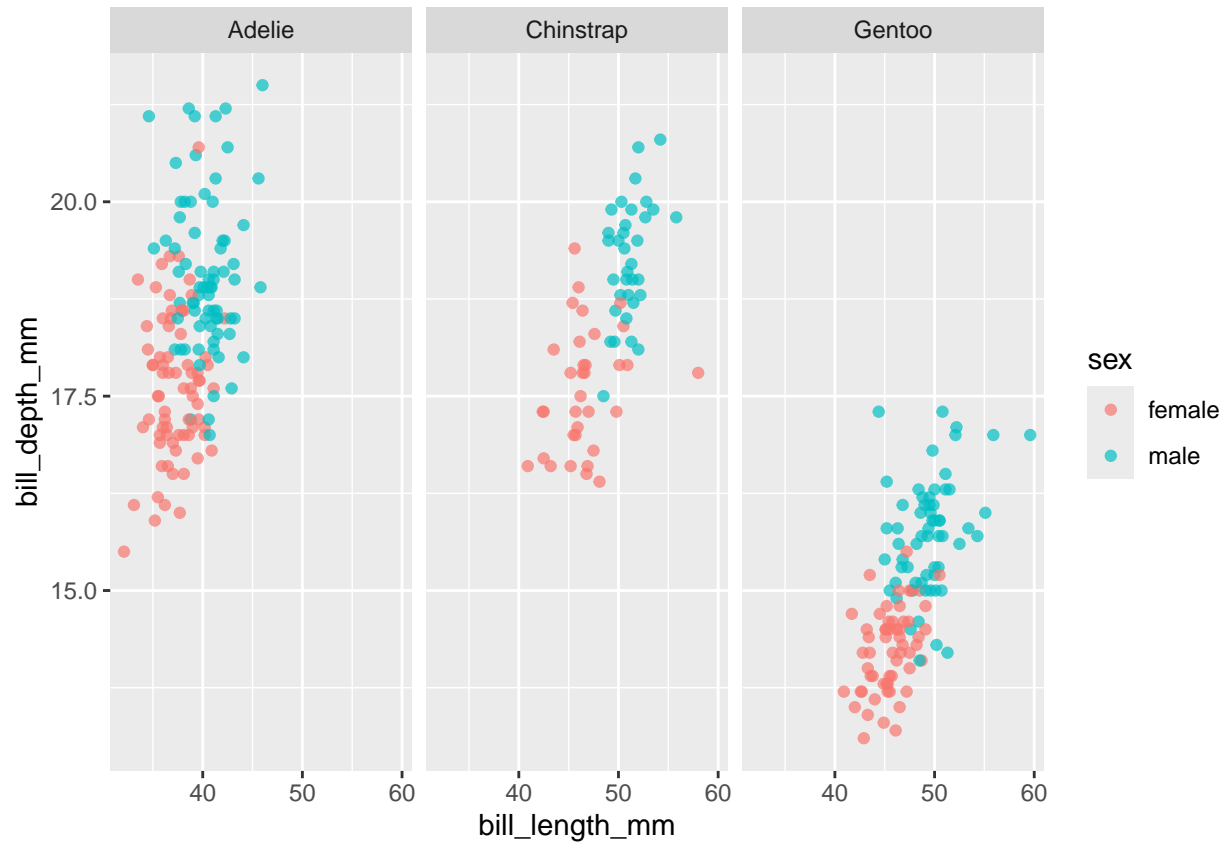


Question 4 Faceting works best with a categorical (discrete) variable cuz you can use it on a quantitative variable but you might get too many panels

Question 5

(Multiple Answers)

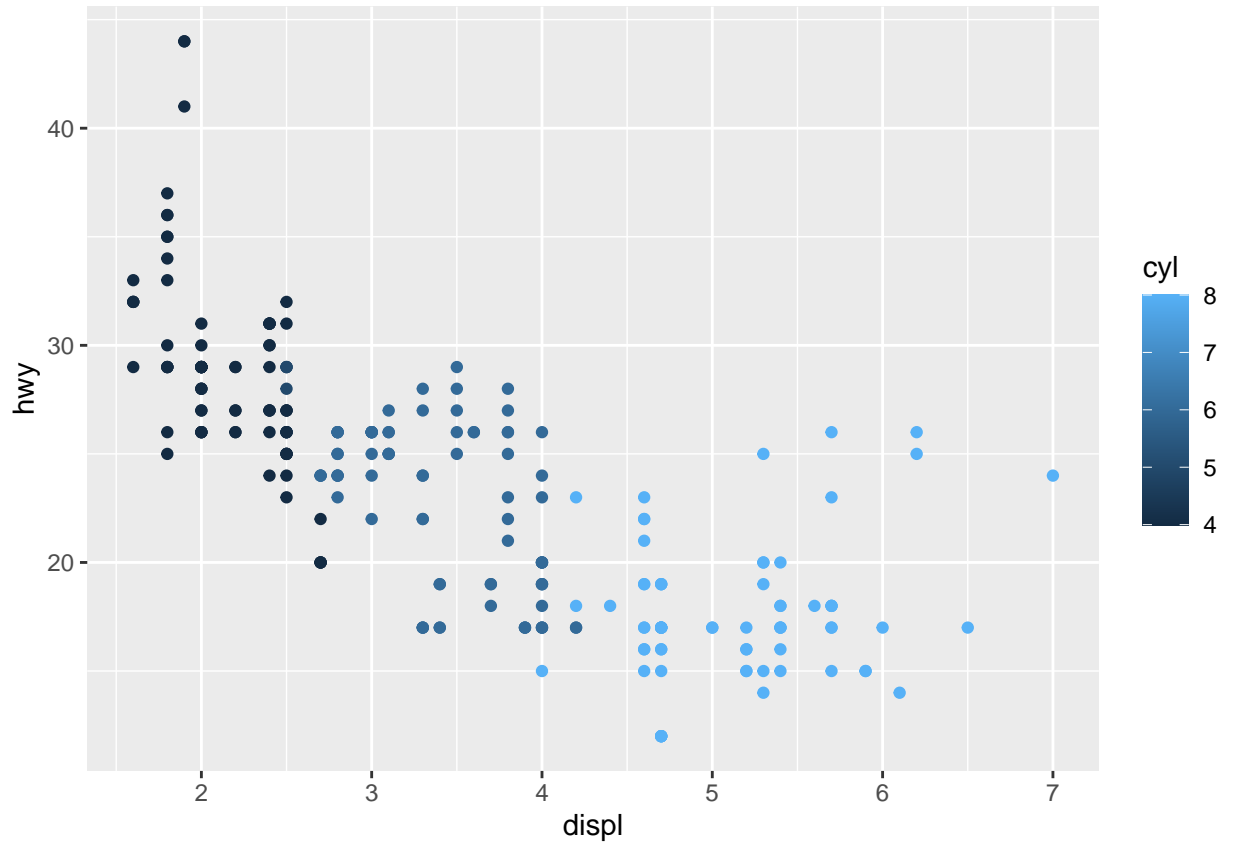
```
ggplot(pengy, aes(x = bill_length_mm, y = bill_depth_mm, color = sex)) +
  geom_point(alpha = 0.7) +
  facet_wrap(~species)
```



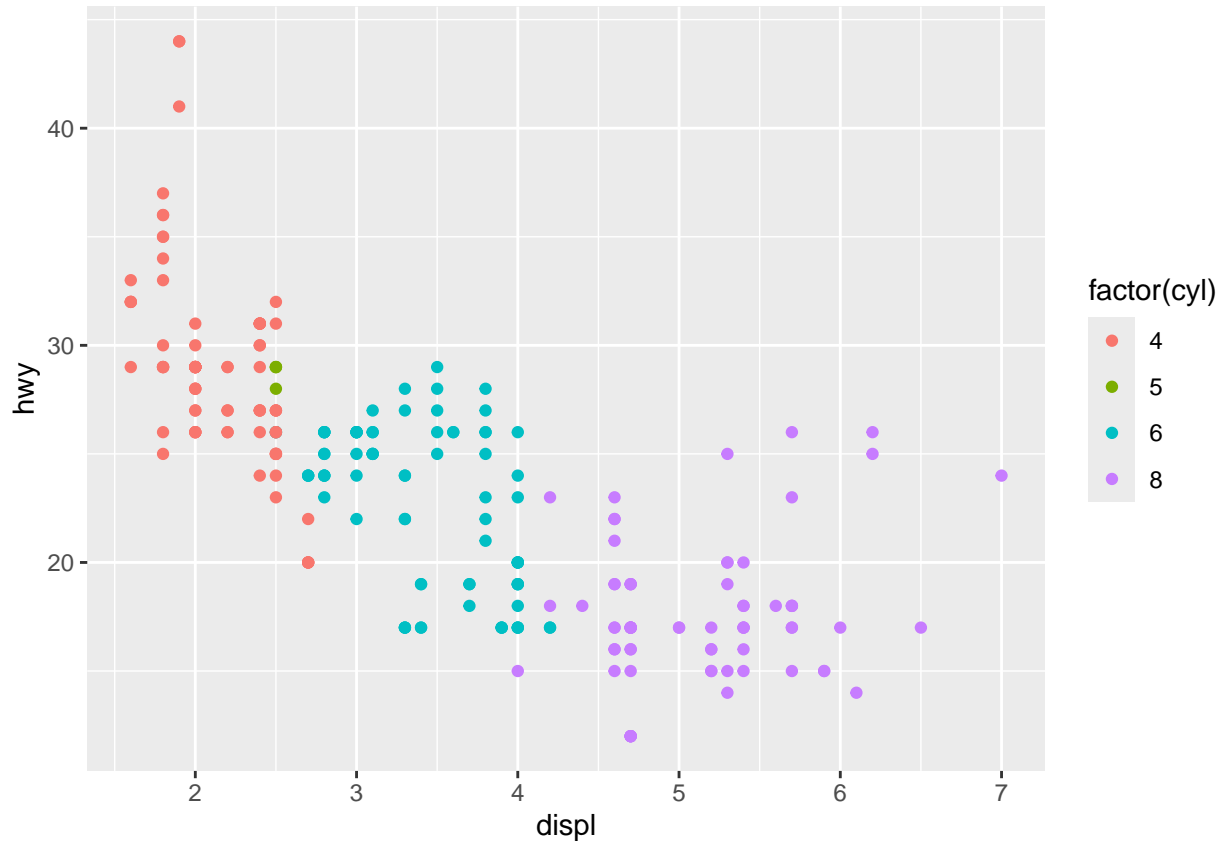
This plot shows the relationship between bill length and bill depth, separated into panels by penguin species and colored by sex.

Question 6

```
# Treat cyl as numeric
ggplot(mpg, aes(displ, hwy, color = cyl)) +
  geom_point()
```



```
# Treat cyl as categorical  
ggplot(mpg, aes(displ, hwy, color = factor(cyl))) +  
  geom_point()
```



Continuous is useful if you want to emphasize ordering (e.g., more cylinders = “more intense” color).

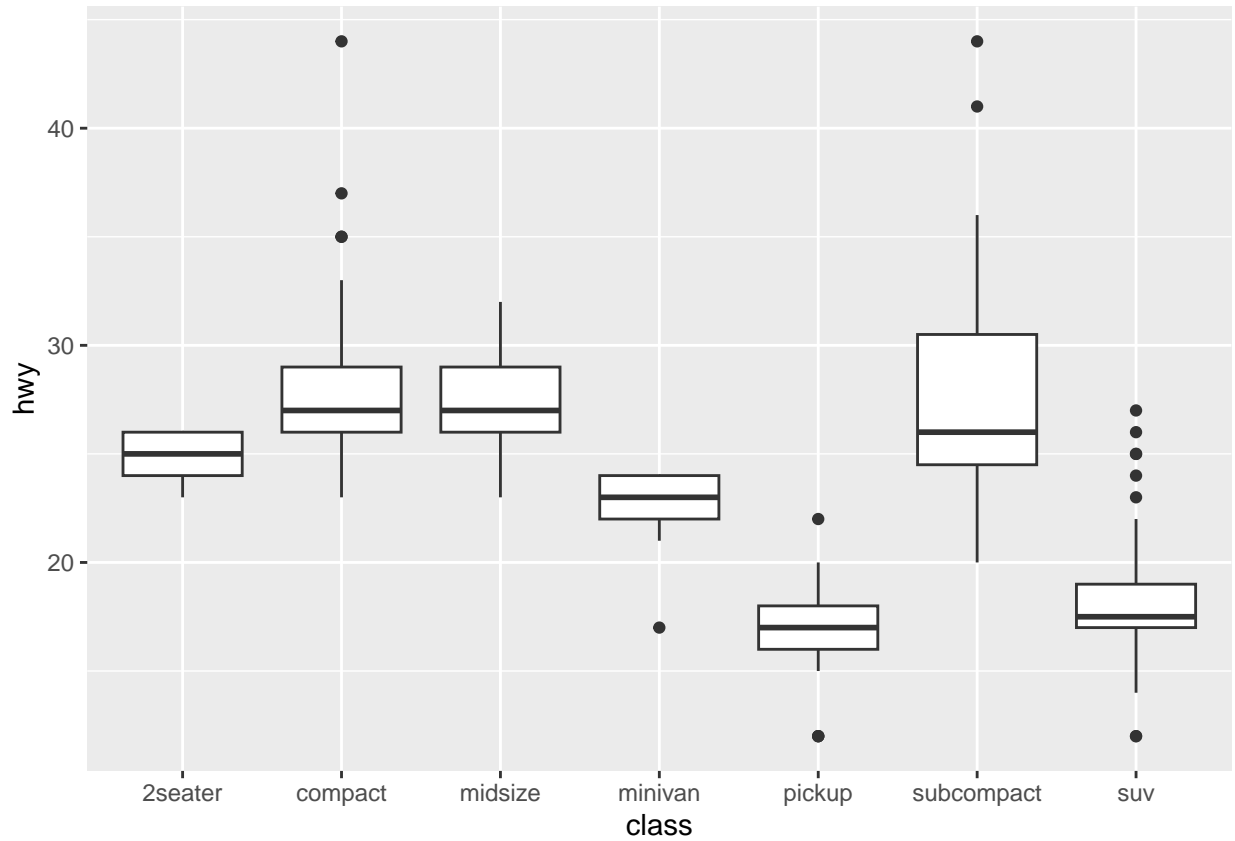
Discrete color is useful if you want clear group comparisons for each cylinder category (4 vs 6 vs 8 vs 5, etc.)

Here, `factor(cyl)` is better cuz it represents count-based groups.

Question 7

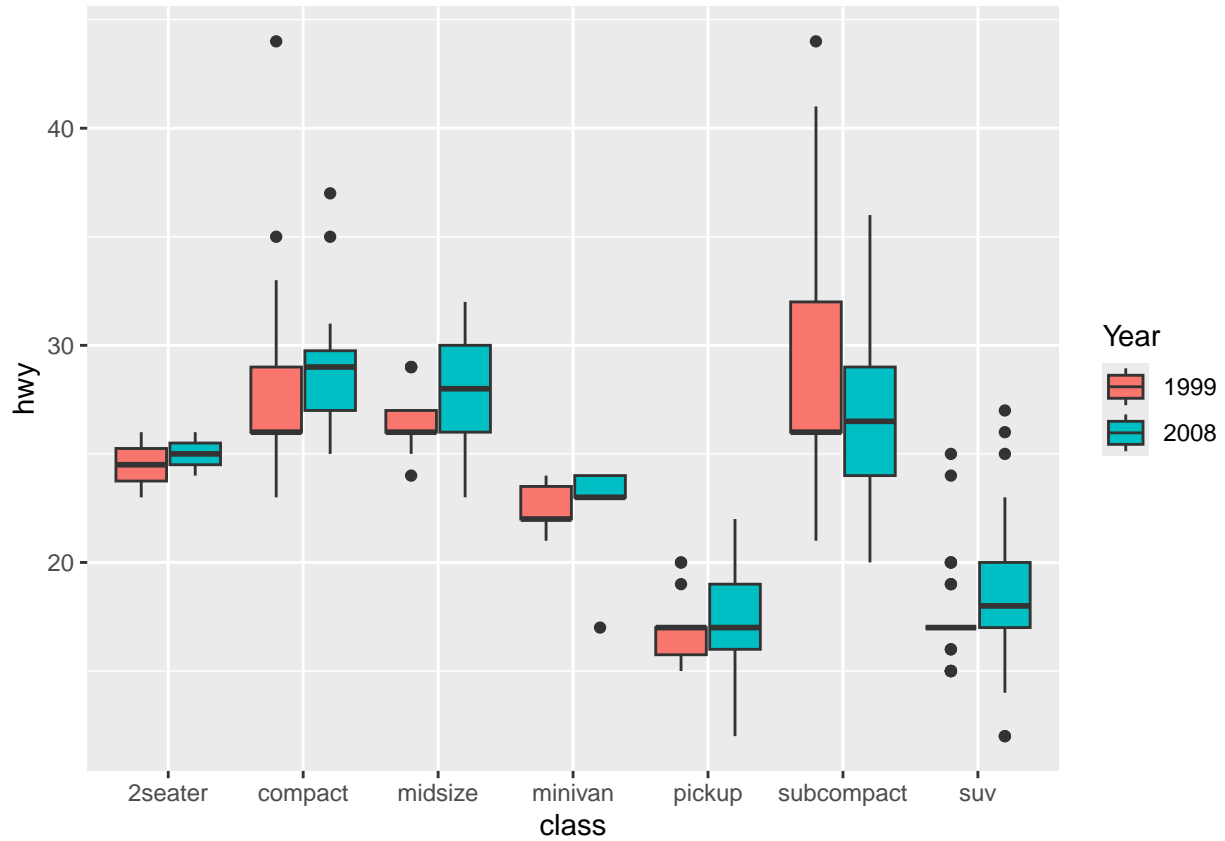
```
ggplot(mpg, aes(x = class, y = hwy, color = year)) +
  geom_boxplot()
```

```
## Warning: The following aesthetics were dropped during statistical transformation:
## colour.
## i This can happen when ggplot fails to infer the correct grouping structure in
## the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
## variable into a factor?
```



Not expected cuz I want two side-by-side boxplots per class (1999 vs 2008). Fix:

```
ggplot(mpg, aes(x = class, y = hwy, fill = factor(year))) +  
  geom_boxplot(position = position_dodge(width = 0.8)) +  
  labs(fill = "Year")
```

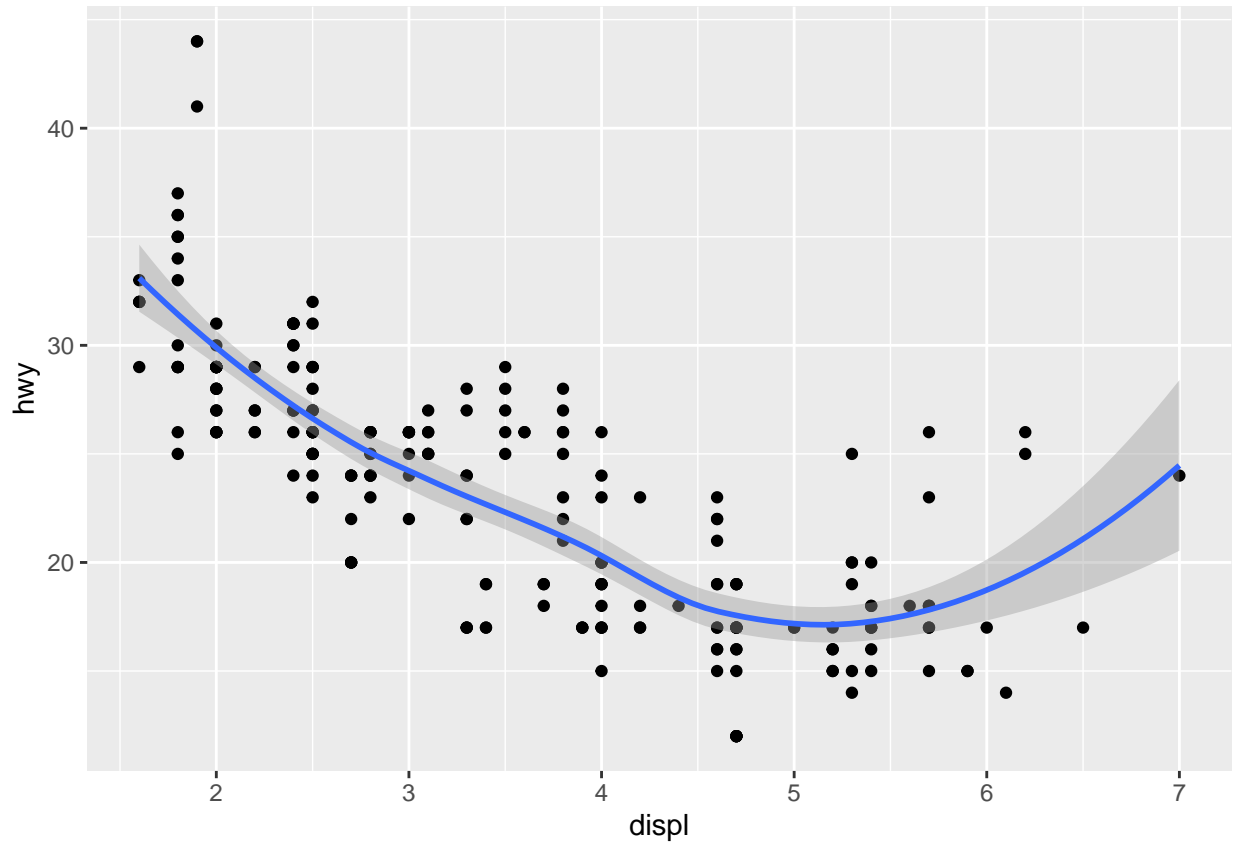


Question 8

Doesn't work cuz `geom_smooth()` is trying to fit a smoother but it doesn't know what x and y are. Fix by defining x and y.

```
ggplot(mpg, aes(displ, hwy)) +
  geom_point() +
  geom_smooth()
```

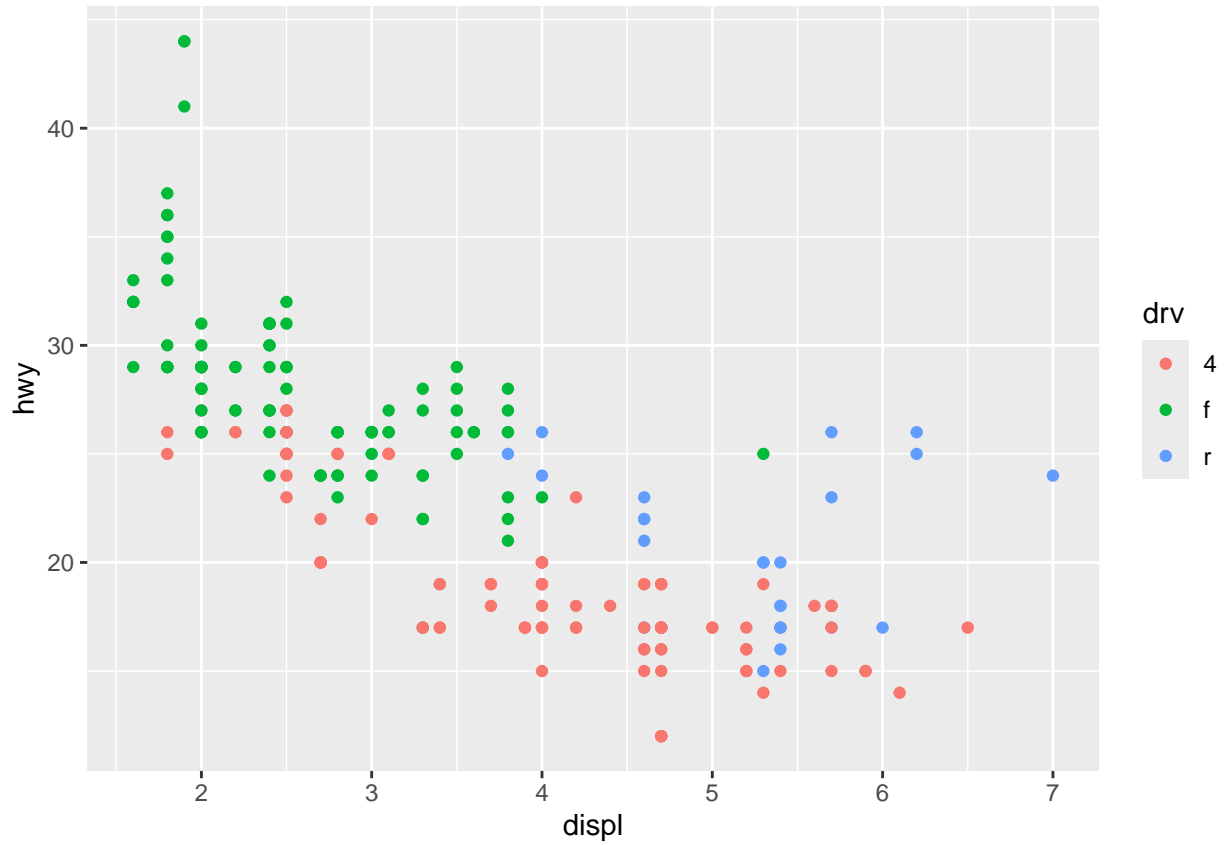
```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



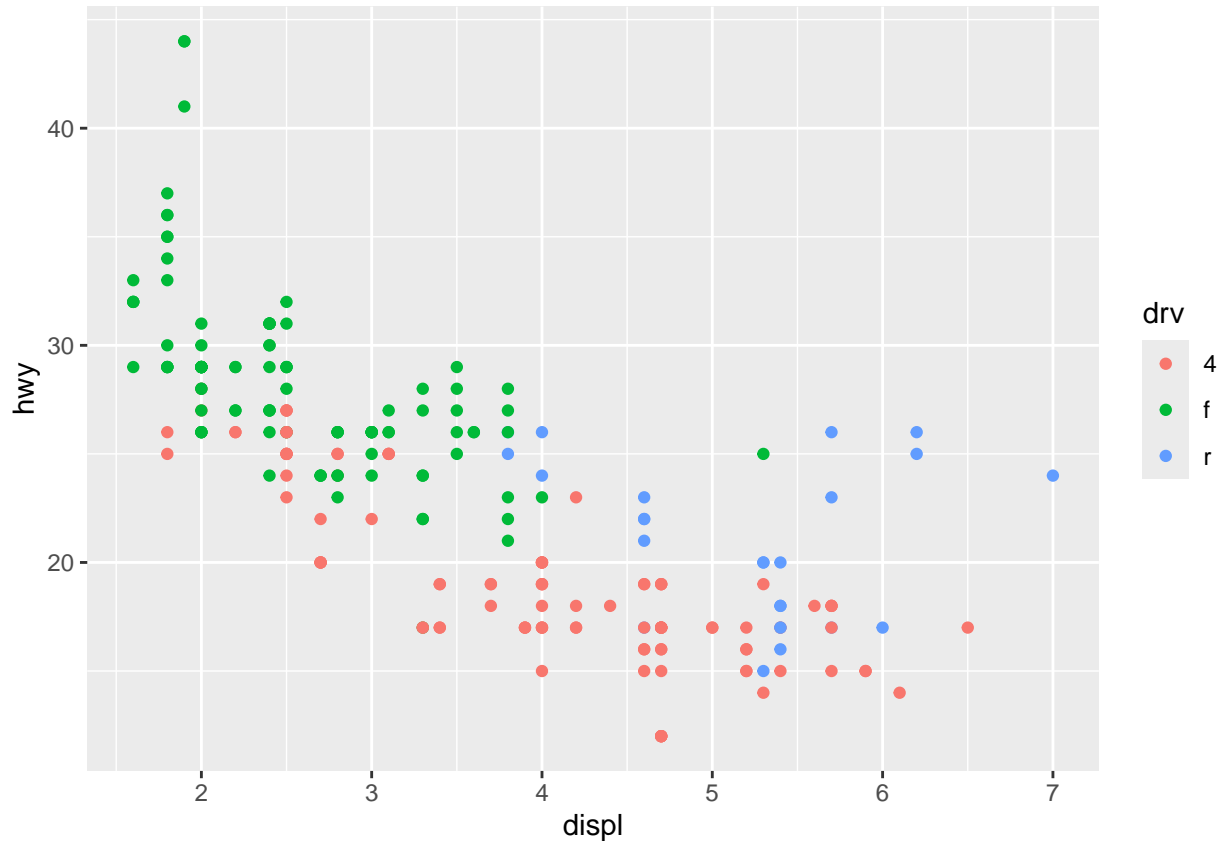
Question 9

How are these different? Even though they look the same

```
ggplot(mpg, aes(displ, hwy)) +  
  geom_point(aes(color = drv))
```



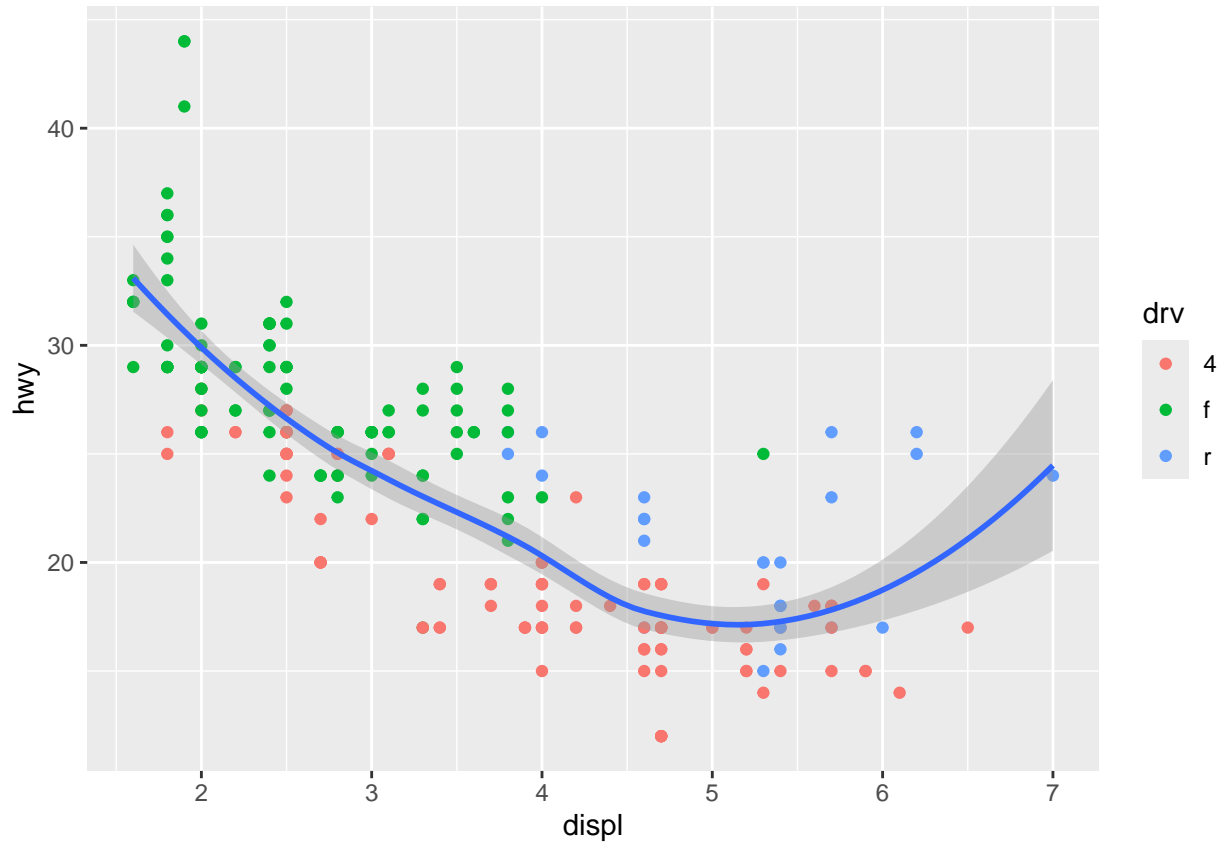
```
ggplot(mpg, aes(displ, hwy, color = drv)) +  
  geom_point()
```



They look the same for now but it will start to look different if you add more layers. for example, `color = drv` in the global will make all layers inherit that color. `color = drv` inside `geom_point()` will make only the points use that grouping. It matters once you add more layers:

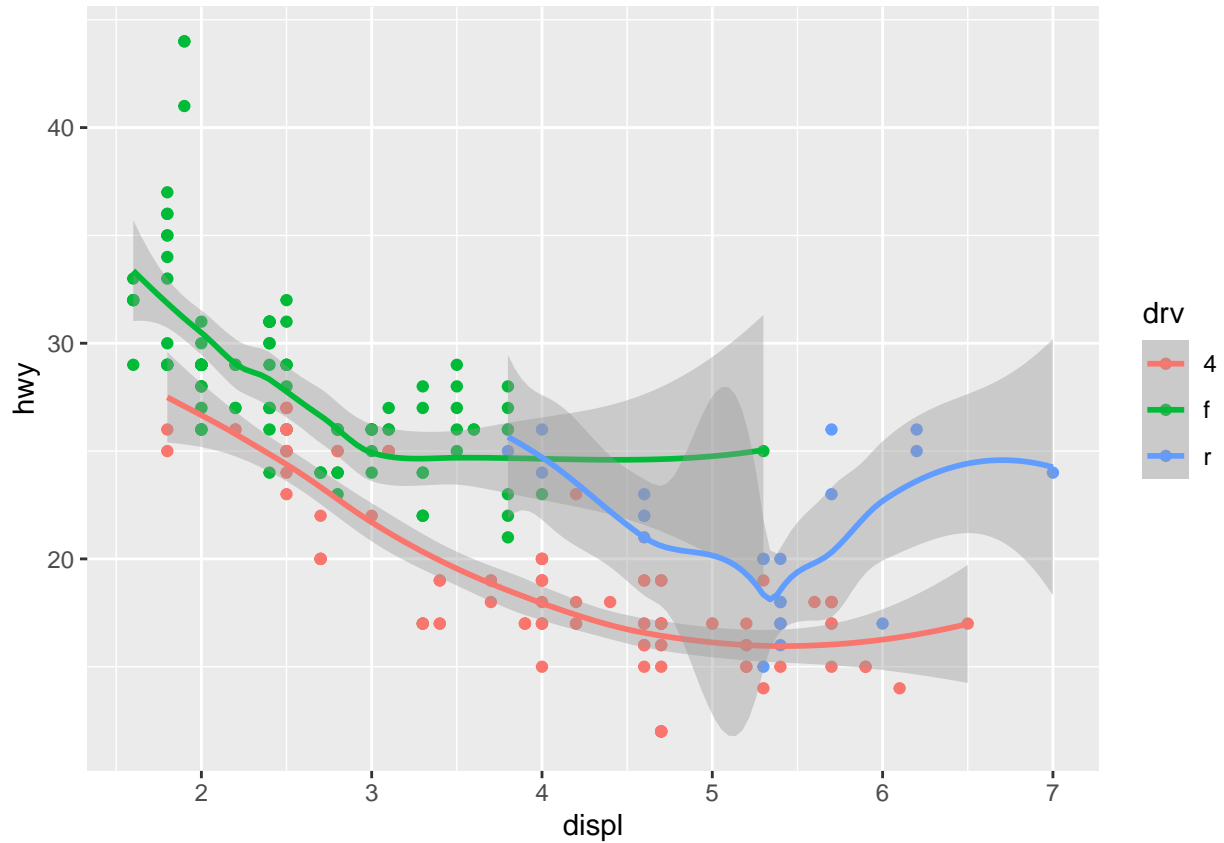
```
## Plot 1
ggplot(mpg, aes(displ, hwy)) +
  geom_point(aes(color = drv)) +
  geom_smooth()
```

`geom_smooth()` using `method = 'loess'` and `formula = 'y ~ x'`



```
## Plot 2
ggplot(mpg, aes(displ, hwy, color = drv)) +
  geom_point() +
  geom_smooth()
```

`geom_smooth()` using method = 'loess' and formula = 'y ~ x'

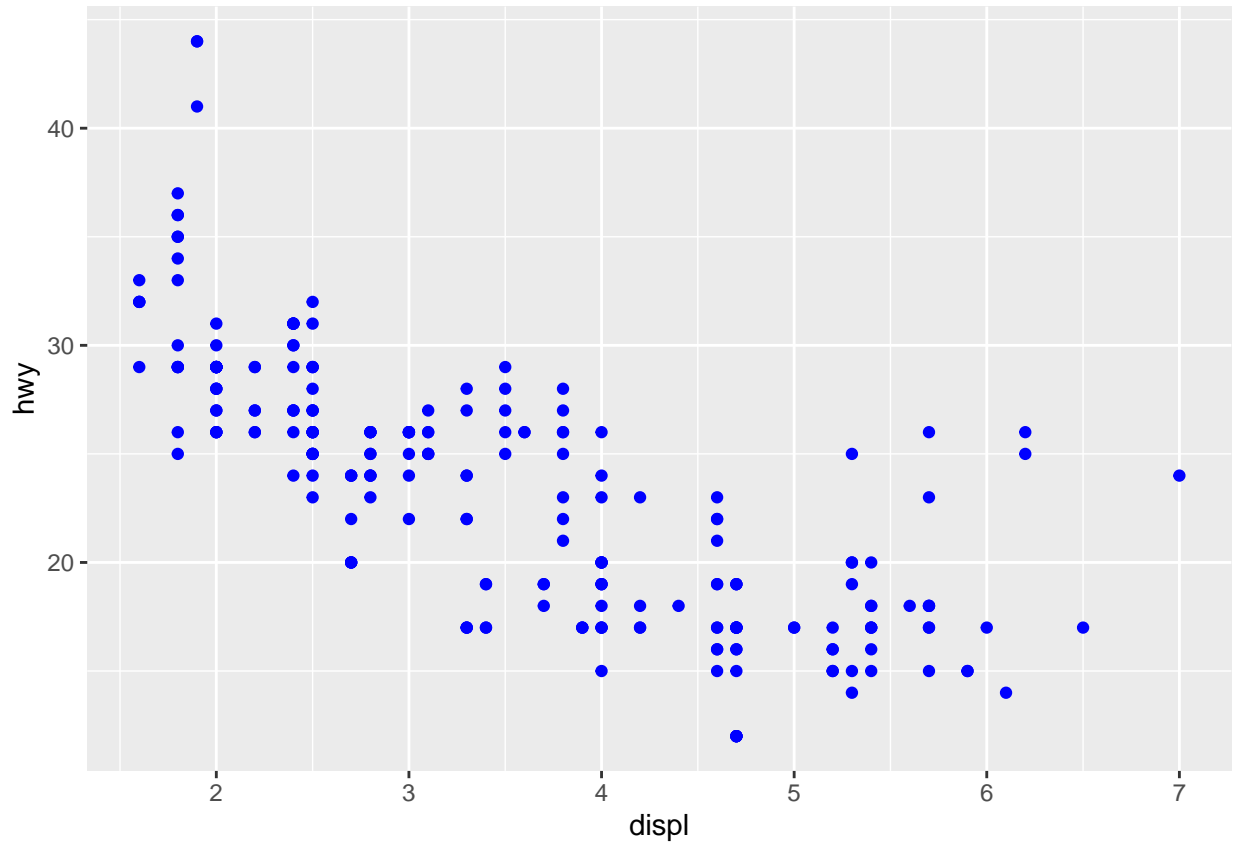


Plot 1: points are colored by `drv`, but `geom_smooth()` is NOT grouped by `drv` so you get one overall smooth line across all data, drawn on top of the colored points.

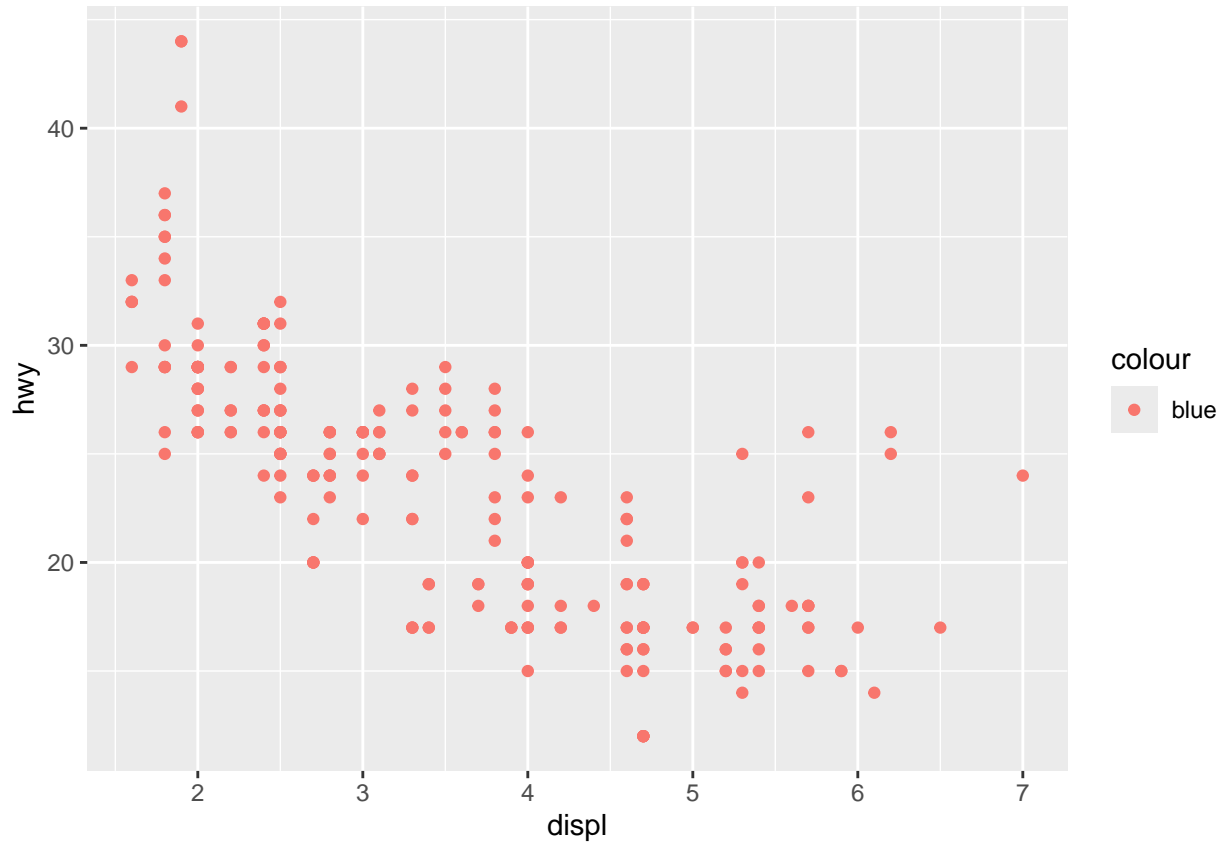
Plot 2: both points and the smoother inherit `color = drv`, so the smoother fits separate smooth lines for each `drv` group so you get multiple smooth curves.

Question 10

```
## Plot 1
ggplot(mpg, aes(displ, hwy)) +
  geom_point(color = "blue")
```



```
## Plot 2  
ggplot(mpg, aes(displ, hwy)) +  
  geom_point(aes(color = "blue"))
```



1. cuz `geom_point(color = "blue")` 2. calls the legend blue but uses default color cuz it hasn't been defined.
3. `aes()` defines mappings from data variables to visual properties